Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
Факультет прикладних інформаційних технологій та електроінженерії
(назва факультету)
Автоматизації технологічних процесів і виробництв
(повна назва кафедри)

ПОЯСНЮВАЛЬНА ЗАПИСКА

до кваліфікаційної роботи

Магістр
(освітній ступінь)

на тему:

Розробка автоматизованої системи розпізнавання мови з з використанням прихованих марківських моделей та нейронних мереж

Виконав: студент (ка) VI курсу, групи КАм-61
Спеціальності 151
“Автоматизація та комп’ютерно-інтегровані технології”
(шифр і назва спеціальності (напряму підготовки))

(підпис) Таран С.А. (прізвище та ініціали)

Керівник (підпис) Дмитрів О.Р. (прізвище та ініціали)

Нормоконтроль (підпис) Козбур І.Р. (прізвище та ініціали)

Завідувач кафедри (підпис) Савків В.Б. (прізвище та ініціали)

Рецензент (підпис) Левицький В.В. (прізвище та ініціали)

м. Тернопіль – 2023

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)

Факультет прикладних інформаційних технологій та електроінженерії

Кафедра автоматизації технологічних процесів і виробництв

Освітній ступінь *Магістр*

Спеціальність *151 "Автоматизація та комп'ютерно-інтегровані технології"*
(шифр і назва)

	ЗАТВЕРДЖУЮ
	Завідувач кафедри

Автоматизації технологічних процесів і виробництв

_____ *Савків В.Б.*

«_____» _____

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ) СТУДЕНТУ

Тарану Сергію Андрійовичу
(прізвище, ім'я, по батькові)

1. Тема проекту (роботи) Розробка автоматизованої системи розпізнавання мови з використанням прихованих марківських моделей та нейронних мереж

Керівник проекту (роботи) Дмитрів Олена Романівна, канд. техн. наук, доцент кафедри АВ
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом по університету від «23» листопада 2023 року № 4/7-1091

2. Термін подання студентом проекту (роботи)

3. Вихідні дані до проекту (роботи) наукові літературні джерела

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Характеристика сучасних систем розпізнавання мови
2. Системи розпізнавання мови, що створені на гібридних моделях, які поєднують в собі нейронні мережі та приховані марківські моделі
3. Створення системи розпізнавання мовних образів
4. Безпека життєдіяльності та основи охорони праці.

5.Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Тема роботи. 2. Актуальність. 3. Мета, об'єкт, предмет, методи та джерела дослідження. 4.Наукова новизна отриманих результатів. 5. Практичне значення. 6. Проблеми розробок систем розпізнавання мови. 7. Найбільш відомі системи розпізнавання мови для портативних комп'ютерів. 8. Класифікація систем розпізнавання мови. 9. Архітектура систем розпізнавання мови. 10. Використання прихованих марківських моделей для розпізнавання мови. 11. Використання нейронних мереж в системах розпізнавання мови. 12. Переваги інтеграції марківських моделей та нейронних мереж при розробці мови. 13. Функціональна схема розпізнавання мови. 14. Структура процесу навчання нейронної мережі. 15. Аглоритм навчання з вчителем. 16. Навчання гібридної моделі 17. Висновки.

6. Консультанти розділів проекту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці та безпека в надзвичайних ситуаціях			
Спеціальна частина	Дмитрів О.Р., канд. техн. наук, доц.		
Нормоконтроль	Козбур І.Р, ст. викл.		

7. Дата видачі завдання *«23» листопада 2023 р.*

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Термін виконання етапів проекту (роботи)	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	23.11 – 23.11	*Виконано*
2.	Підбір джерел за темою кваліфікаційної роботи	24.11 – 26.11	*Виконано*
3.	Опрацювання джерел за темою кваліфікаційної роботи	27.11 – 30.11	*Виконано*
4.	Виконання дослідження щодо розробки методів пошуку рішень с системах розпізнавання мови, що ґрунтуються на гібридних моделях	01.12 – 06.12	*Виконано*
5	Розробка алгоритмів	07.12 – 10.12	*Виконано*
6.	Оформлення розділу «Огляд існуючих систем розпізнавання мови»	11.12 – 13.12	*Виконано*
7.	Оформлення розділу «Системи розпізнавання мови на основі гібридних моделей нейронних мереж та прихованих харківських моделей»	14.12 – 15.12	*Виконано*
8.	Оформлення розділу «Побудова системи розпізнавання мови»	16.12 – 18.12	*Виконано*
9.	Оформлення розділу «Охорона праці та безпека в надзвичайних ситуаціях»	06.12 – 16.12	*Виконано*
10.	Оформлення кваліфікаційної роботи	14.12 – 19.12	*Виконано*
11.	Нормоконтроль	19.12 – 20.12	*Виконано*
12.	Перевірка на плагіат	20.12 – 25.12	*Виконано*
13.	Попередній захист кваліфікаційної роботи	25.12 – 27.12	*Виконано*
14.	Захист кваліфікаційної роботи	29.12	

Студент _____ Таран С.А.
(підпис) (прізвище та ініціали)

Керівник проекту (роботи) _____ Дмитрів О.Р.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Розробка автоматизованої системи розпізнавання мови з використанням прихованих марківських моделей та нейронних мереж // Таран Сергій Андрійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет прикладних інформаційних технологій та електроінженерії, кафедра автоматизації технологічних процесів і виробництв, група КАм-61 // Тернопіль, 2023 // с. – 90, слайд. – 17, рис. – 20, табл. – 4 , бібліогр. – 22.

Ключові слова: ПРИХОВАНА МАРКІВСЬКА МОДЕЛЬ, НЕЙРОННА МЕРЕЖА, СИСТЕМА, ПЕРЦЕПТРОН, АЛГОРИТМ, ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ.

Кваліфікаційна робота присвячена питанням вирішення проблем розробки нових систем розпізнавання мови. Розглядаються існуючі системи розпізнавання мовних образів, проведено їх аналіз позитивних і негативних характеристик. Запропонований проект розробки автоматизованої системи розпізнавання мови з використанням гібридних моделей комбінованої мови - прихованих марківських моделей та нейронних мереж.

У вступі висвітлені актуальність та обґрунтування досліджуваного питання. Перший розділ містить аналіз систем класифікації мовного розпізнавання та математичне моделювання мовних сигналів. Другий розділ розглядає математичні моделі процесів мовоутворення та сприйняття, а також ефективність систем на різних рівнях абстракції. Третій розділ описує технологічний підхід, використовуючи приховані марківські моделі та нейронні мережі, зокрема TDNN і RNN. У четвертому розділі представлено конструкторську частину, включаючи гібридні моделі. П'ятий розділ досліджує спеціальну частину, охоплюючи процес створення системи та навчання нейронних мереж.

Робота пропонує інноваційний підхід до розпізнавання мови, що може визначити нові стандарти та стимулювати подальше розвиток галузі.

ANNOTATION

Development of an automated speech recognition system using hidden Markov models and neural networks // Serhiy Taran// Ternopil Ivan Pul'uj National Technical University, Faculty of Applied Information Technologies and Electrical Engineering, Department of Automation of Technological Processes and Productions // Ternopil, 2023 // p. – 90, Fig. - 20, Table – 4, Slides – 17, References – 22.

Key words: HIDDEN MARKOV MODEL, NEURAL NETWORK, SYSTEM, PERCEPTRON, ALGORITHM, SOFTWARE.

The qualification work is dedicated to addressing the issues of developing new language recognition systems. Existing speech recognition systems are examined, and their positive and negative characteristics are analyzed. The proposed project involves the development of an automated language recognition system using hybrid models of combined language, specifically hidden Markov models and neural networks.

In the introduction, the relevance and justification of the research question are highlighted. The first chapter includes an analysis of speech recognition classification systems and mathematical modeling of speech signals. The second chapter examines mathematical models of language formation and perception processes, as well as the efficiency of systems at different levels of abstraction. The third chapter describes the technological approach using hidden Markov models and neural networks, including TDNN and RNN. The fourth chapter presents the design part, including hybrid models. The fifth chapter explores the specific aspect, covering the system creation process and training of neural networks.

The work proposes an innovative approach to speech recognition that could set new standards and stimulate further development in the field.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ

АРМ – автоматичне розпізнавання мови.

БП – багатошаровий персептрон.

ВР – алгоритм зворотного поширення помилки

ПММ – приховані марківські моделі.

СРМ – системи розпізнавання мови.

ШНМ – штучні нейронні мережі.

TDNN (Time-Delay Neural Network) – нейронні мережі із затримкою часу.

RNN (Recurrent Neural Network) – рекурентні нейронні мережі.

ЗМІСТ

ВСТУП

Актуальність теми. Автоматизовані системи розпізнавання мови (АРМ) є частиною обробки мови та призначені для полегшення взаємодії між користувачем і машиною. Вони можуть використовуватися для різних завдань, від розпізнавання простих команд до складних систем розпізнавання природньої мови. Системи АРМ вбудовуються в різні додатки, такі як системи голосового контролю, навчання мови та інші.

Ці системи корисні для пошуку та сортування аудіо- та відеоданих та використовуються для введення інформації, коли користувач не може використовувати очі або руки, наприклад, на робочому місці чи за кермом автомобіля. Системи АРМ дозволяють працюючим в напруженій обстановці використовувати комп'ютер для отримання або введення необхідної інформації.

Зазвичай системи АРМ застосовуються в таких системах, як телефонні додатки, вбудовані системи (системи набору номера, робота з кишеньковим комп'ютером, керування автомобілем і т.п.), мультимедійні додатки (системи навчання мови) та інші.

Розробки в області розпізнавання мовлення почалися ще в 1920-х роках, але реальний прогрес став помітний тільки у 1952 році з створенням компанією Bell Laboratories першої такої системи. Сьогодні в цьому напрямку працюють вже не десятки, а сотні дослідницьких колективів у наукових та навчальних закладах, а також у великих корпораціях. Про це можна судити з таких міжнародних форумів вчених та фахівців в області мовних технологій як ICASSP, EuroSpeech, ICPHS та ін. Розробки в області розпізнавання мовлення включають різні технології, в тому числі приховані марківські моделі (ПММ) і штучні нейронні мережі (ШНМ).

Ключовими труднощами при розробці систем автоматичного розпізнавання мови (АРМ) є велика різноманітність вимови та вплив різноманітних факторів, таких як шум та спотворення, на вхідний сигнал. Ряд факторів, таких як оточуючий шум, відбиття, ехо та перешкоди в каналі, впливають на якість сигналу. Складність збільшується тим, що характеристики шуму та спотворень заздалегідь невідомі, що

ускладнює можливість попереднього налаштування системи на ці параметри перед її використанням.

Сучасні технології, що широко використовуються сьогодні, ґрунтуються на прихованих марківських моделях та штучних нейронних мережах. Обидва підходи мають свої плюси та мінуси, науковці активно працюють над розробкою гібридних моделей, які зможуть поєднати переваги їх обох.

Об'єкт дослідження. Процес автоматизованого розпізнавання мовних образів.

Предмет дослідження. Гібридні моделі мови, що комбінують приховані марківські моделі та нейронні мережі.

Методи дослідження. Вирішення завдань розпізнавання мови за допомогою підходів, що базуються на методах теорії ймовірностей, теорії випадкових процесів та теорії нейронних мереж.

Практичне значення одержаних результатів. Отримані результати дослідження та розробки автоматизованої системи розпізнавання мови з використанням прихованих марківських моделей та нейронних мереж мають значущі практичні застосування в різних сферах. Нижче наведено деякі з практичних вигод та можливостей, які можуть виникнути з використання отриманих результатів:

1. Підвищення точності розпізнавання мови**:** Запропонована система, комбінуючи приховані марківські моделі та нейронні мережі, може досягти високої точності розпізнавання мови навіть в умовах шуму чи інших спотворень. Це робить систему ефективною в реальних сценаріях використання, таких як розпізнавання мови в шумному оточенні або в системах голосового керування.

2. Широкі можливості застосування в індустрії та бізнесі**:** Система може бути успішно використана в індустріальних та бізнесових сценаріях, зокрема для автоматизованого оброблення голосових команд, створення диктантів або автоматизованого аналізу телефонних розмов. Це може покращити ефективність роботи та взаємодії з інформацією.

3. Розвиток систем голосового інтерфейсу**:** Результати дослідження можуть сприяти подальшому розвитку голосових інтерфейсів в різних пристроях,

таких як мобільні телефони, планшети, домашні асистенти та інші. Висока точність розпізнавання забезпечить зручність та ефективність користувачів.

4. Медичні застосування**:** В сфері медицини система може бути використана для розпізнавання та документування мовленнєвих патологій, а також для розвитку інтерфейсів для людей з обмеженими можливостями.

5. Розширення мовних можливостей інтернет-платформ**:** Застосування системи в інтернет-сервісах, таких як пошукові системи, асистенти та інші, може покращити розпізнавання голосових запитань та команд, забезпечуючи більш швидку та точну взаємодію з користувачами.

Наукова новизна одержаних результатів.

1. Проведено аналіз розробки голосових інтерфейсів, що ґрунтується на механізмі розпізнавання фраз користувача.

2. Запропоновано вибір для гібридної моделі нейронні мережі з рекурентною архітектурою та із затримкою часу.

3. Розглянута гібридна модель комбінованої мови, яка представляє собою синтез прихованих марківських моделей та штучних нейронних мереж, що дозволяє ефективно поєднати переваги прихованих марківських моделей із можливостями штучних нейронних мереж.

4. Запропонований алгоритм Вітербі пошуку параметрів прихованих марківських моделей при розв'язанні задач автоматизованого розпізнавання мови на основі гібридної моделі.

5. Розроблена функціональна схема системи автоматизованого розпізнавання мови.

6. Розглянуто алгоритм навчання для запропонованих нейронних мереж, а саме: алгоритм Больцмана для навчання рекурентних мереж та алгоритм зворотного поширення похибки для мереж із затримкою часу.

Особистий внесок. Усі результати, які становлять основний зміст магістерської кваліфікаційної роботи, автор отримав особисто, а саме шляхом проведення ґрунтовного огляду літературних джерел за темою кваліфікаційної

роботи, проведена систематизація та аналіз передового наукового досвіду в галузі мовного розпізнавання, а також впровадження нових ідей та підходів у даному контексті, розглянуто та оцінено переваги та обмеження прихованих марківських моделей та нейронних мереж у контексті розпізнавання мови, розроблено та вдосконалено алгоритми обробки сигналів для підвищення якості вхідних даних, розроблено та оптимізовано програмне забезпечення для реалізації запропонованої системи.

Усі ці складові мого внеску об'єднуються з метою створення автоматизованої системи розпізнавання мови, яка відповідає найвищим стандартам точності та продуктивності, що сприяє не лише практичному вдосконаленню технічних аспектів системи, але й вносить свій внесок у розвиток області мовного розпізнавання в цілому.

Апробація результатів роботи. Окремі результати роботи доповідалися на XI Всеукраїнській студентській науково-технічній конференції «Інформаційні моделі, системи та технології», що проводилася 13-14 грудня 2023 року у Тернопільському національному технічному університеті імені Івана Пулюя [20].

Структура роботи. Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 6 розділів, висновків та переліку використаних джерел. Обсяг роботи: розрахунково-пояснювальна записка – 90 арк. формату А4, графічна частина – 17 графічних слайдів.

РОЗДІЛ 1
АНАЛІТИЧНА ЧАСТИНА

1.1 Системи класифікації мовного розпізнавання: огляд та аналіз

У зв'язку з успіхами розвитку обчислювальної техніки та нових інформаційних технологій, в останнє десятиліття визначилася тенденція до зростання складності керуючих систем, а також усіх інших видів «людино-машинного» управління. Важливою є можливість взаємодії людини з машиною мовою, максимально наближеною до природної мови людини. Застосування розпізнавання мови в керуючих системах як інтерфейс взаємодії «людина-машина» дозволяє організувати ефективну та зручну взаємодію користувача з системою.

Відомо, що мова включає кілька видів інформації. Основний вид інформації – семантична, яка передає зміст повідомлення, його суть. Однак даним видом інформації роль мови у спілкуванні людей не вичерпується: велике значення має емоційне забарвлення мови (інтонація).

З погляду акустичної теорії мова є акустичний сигнал, який можна розділити на смислові одиниці – слова. Набір слів є фразою, при цьому слова складаються з окремих звуків. Для завдання розпізнавання важливими є лише основні, самостійні звуки, що відрізняють слова один від одного, які називаються фонемами. Фонема не є поодиноким звуком і може бути представлена як серія звуків зі схожими характеристиками, які називаються алофонами. Залежно від розташування фонем по сусідству вони звучать по-різному через вплив. Цей ефект називають коартикуляцією. З вищеназваної причини деякі системи розпізнавання мовлення працюють не з фонемами, а з складнішими звуковими одиницями: дифонами, трифонами.

Класична функціональна схема системи розпізнавання мовної інформації складається з наступних функціональних вузлів:

- мікрофон;
- блок обробки;

– блок аналізу.

Важливо відзначити, що з урахуванням програмного та апаратного забезпечення внутрішній пристрій блоків системи розпізнавання мови набагато складніший.

Мова у вигляді звукових хвиль фіксується мікрофоном, який перетворює їх на аналоговий мовний сигнал.

У блоці обробки аналоговий сигнал перетворюється на цифрову форму, проходить фільтрацію та попередню корекцію, розбивається на ділянки, в яких відбувається виділення акустичних параметрів для подальшого аналізу. Блок аналізу зазвичай включає акустичний, лінгвістичний та семантичний аналізи. Аналізуються набори акустичних параметрів зі звуковими образами слів як еталонів чи моделей, сукупність яких називається словником. Зазвичай словник створюється на етапі розробки системи і може доповнюватися та коригуватися в подальшому під час експлуатації під конкретного користувача. Процес створення еталонів часто проходить в інтерактивному режимі і зветься навчанням системи.

Розпізнавання окремих мовних команд простіше, ніж розпізнавання злитої мови і вимагає великих обчислювальних потужностей. Саме з цієї причини на сьогоднішній день існує величезний вибір програмного та апаратного забезпечення, що має невелику вартість та помірну якість розпізнавання. Проте, тести показують, що системи так і не подолали рівень розпізнавання в 80 %, тоді як у людини цей показник становить 96–98 %.

Щоб дати повну оцінку сучасному стану систем розпізнавання мови, автором представлена класифікація систем за такими основними параметрами.

Технічне виконання. Усі системи розпізнавання мови за технічним виконанням можна поділити на програмні продукти та програмно-апаратні засоби. Перші реалізуються у вигляді програмного забезпечення, що інсталюється на комп'ютеризовану техніку, другі є закінченим пристроєм.

Призначення. За цим параметром всі системи розпізнавання поділяються на види:

– командні системи;

– системи диктування;

– системи розпізнавання.

Персоналізація. Під цим параметром ховається залежність (чи незалежність) системи розпізнавання диктора. Усі системи розпізнавання мови поділяються на дикторозалежні та дикторонезалежні системи. Перші призначені для роботи тільки з одним користувачем (система навчена для однієї людини), другі призначені для роботи з будь-яким користувачем.

Тип мови. Мову користувача умовно можна розділити на злиту та роздільну. Якщо слова у мові розділені між собою ділянками тиші, така мова вважається роздільною. До злитого мовлення належать природно вимовлені пропозиції.

За типом мови системи розпізнавання поділяються на системи, що розпізнають роздільну мову, та системи, що розпізнають злиту мову.

Розмір словника. Під розміром словника систем розпізнавання розуміється кількість слів, які система може розпізнати. За цим критерієм вирізняють системи з обмеженим словником та системи зі словником великого розміру.

Тип структурної одиниці. При розпізнаванні мови як структурної одиниці можуть бути обрані окремі слова або частини слів, такі як фонеми, алофони, дифони і трифони. Системи, що використовують при розпізнаванні слова цілком або фрази, називаються системами розпізнавання за зразком. Створення таких систем менш трудомістке на відміну від систем, що розпізнають мінімальні структурні одиниці мови: фонеми, алофони, дифони та трифони. Таким чином, системи розпізнавання на кшталт структурної одиниці поділяються на системи розпізнавання за еталоном та системи розпізнавання за структурною одиницею.

Механізм функціонування. У сучасних системах розпізнавання широко використовуються різні підходи до механізму функціонування, серед яких найбільшу популярність набули такі:

– приховані марківські моделі;

– динамічне програмування;

– нейромережевий метод;

– експертні системи;

– найпростіші детектори.

Узагальнивши все вищезгадане, можна уявити класифікацію систем розпізнавання мови представлену на рис. 1.1.

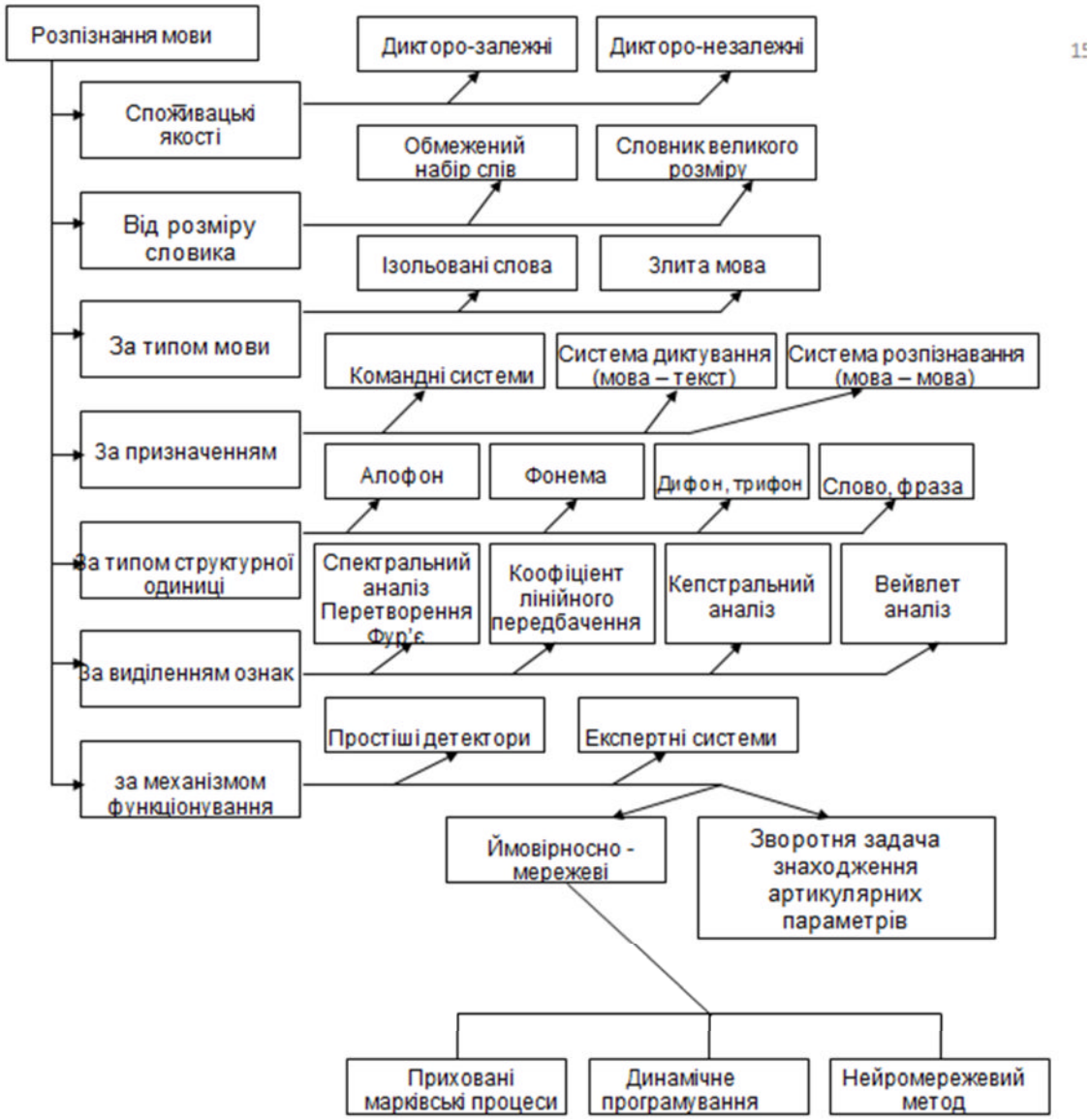

Рис.1.1 Класифікація систем розпізнавання мови

1.2. Математичне моделювання мовних сигналів

У ранніх версіях пристроїв автоматичного розпізнавання мови, розроблених у 1948 році, використовувалася порогова логіка для грубої обробки аналогових напруг, що представляли акустичні образи. Ця обробка визначалася на значних інтервалах часу з метою розпізнавання слів або коротких висловів від одного

диктора. Відразу стало очевидним, що такий простий підхід не є достатнім для ефективного розпізнавання мови.

Всі існуючі на сьогодні системи розпізнавання мови базуються на двох ключових концепціях, прийнятих в більшості лінгвістичних теорій: перша гіпотеза стверджує, що інформація мовним сигналом виражається в часових змінах амплітудного спектру, а друга визнає, що мова є складним ієрархічно організованим сигналом. Лінгвістичні теорії вибирають спектральні або періодичні характеристики сигналу як найпростіші вихідні образи, з яких формуються більш складні лінгвістичні конструкції, такі як фонеми, фонетичні категорії і інші. У даній роботі розглядаються математичні моделі класифікації мовних сигналів, такі як акустичне моделювання, динамічне програмування, статистичні мовні образи, приховані марківські моделі та штучні нейронні мережі.

1.3 Акустичне і динамічне моделювання

Найпростішим методом розпізнавання ізольованих слів є порівняння акустичних сигналів цих слів. Його суть полягає у тому, що окремі фонеми ідентифікуються за їхніми акустичними властивостями, такими як тривалість та головна частота. Акустичні ознаки для машинного розпізнавання мови включають частоти формант, частоту основного тону, антирезонанси, потужність сигналу та тривалість голосних чи приголосних.

Більшість досліджень у цій області використовують стратегію сегментації та фонетичної ідентифікації акустичних сигналів. Системи розпізнавання використовують часове вирівнювання для аналізу випадків, коли еталони слів та мовні сигнали мають різну часову структуру. Рекурсивна формула динамічного вирівнювання використовується для обчислення відстані між послідовностями векторів ознак.

Акустичне і динамічне моделювання мовних сигналів представляє собою інтегрований підхід у сфері АРМ.

Акустичне моделювання зосереджується на аналізі звукових характеристик мовлення, таких як частоти формант, основного тону, антирезонанси та інші акустичні ознаки. Його метою є створення математичних моделей, які визначають, як ці акустичні ознаки пов'язані з мовленнєвими одиницями.

Динамічне моделювання використовує динамічне програмування для вирішення проблем вирівнювання або синхронізації між вхідним акустичним сигналом і відомим еталоном чи моделлю мовлення.

Ранні системи розпізнавання використовували класифікаторну архітектуру, яка зображена на рис. 1.2.

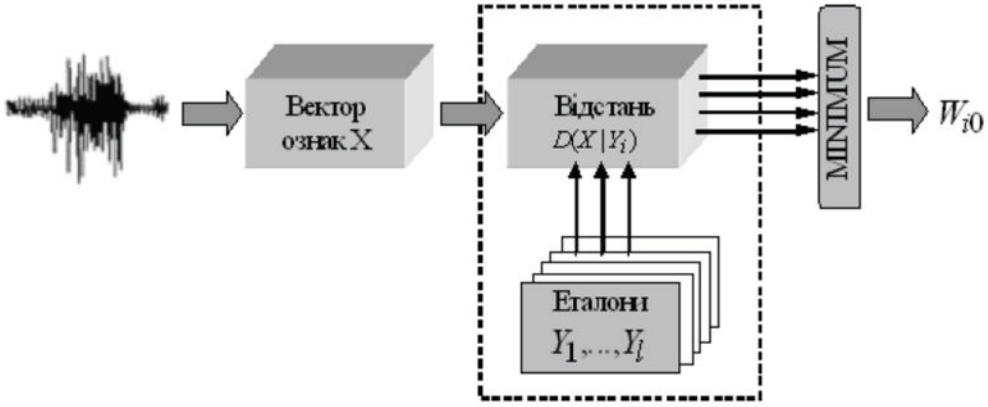

Рис. 1.2. Структура автоматичного класифікатора

Представлена структура автоматичного класифікатора, де кожне слово *Wl* має свій еталон *Yl*. Кожен еталон *Yl* складається з хронології векторів ознак для різних реалізацій слова *Wl*. Для віднесення невідомого слова до одного з класів класифікатора, необхідно здійснити порівняльну характеристику послідовності векторів ознак цього слова з еталонами *Yl,1,...,Yl*. Рішення приймається на користь слова, яке має найкращий відгук еталона *l* W*, використовуючи правило:

$$l^* = \arg\min_{l} \min_{m} D(X, Y_{l,m}). \tag{1.1}$$

де *l** - це відстань між послідовністю входу та послідовністю еталону, які є різними за довжиною. Ця відстань обчислюється як сума локальних відстаней вздовж шляху вирівнювання між послідовностями векторів.

Для обчислення кумулятивної відстані $D_{i,j} = D(x_1,\ldots,x_i,y_1,\ldots,y_j)$ між двома послідовностями *X, Y* векторів використовується рекурсивна формула (1.2).

$$Di,j = \begin{cases} \min\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\} + d_{ij} & i=j=0 \\ \infty & i>0, j>0 \\ 0 & \text{в іншому випадку} \end{cases} \quad (1.2)$$

де $D(X,Y)=D_{ST}$ загальна відстань між векторами, що розраховується за $O(S \cdot T)$ обчислювальних операцій.

Цей продуктивний метод для обчислення відстані між послідовностями, що не залежить від часу, відомий як алгоритм динамічного вирівнювання та активно використовується з 1970 року у різноманітних варіантах автоматизованих систем розпізнавання слів.

1.4 Класифікація мовних образів на основі статистичних методів: огляд та аналіз

Статистичний метод розпізнавання мови є стандартом в галузі АРМ та використовує архітектуру із чотирма основними компонентами, що включають виділення акустичних ознак, акустичну та мовну моделі, процедуру глобального пошуку.

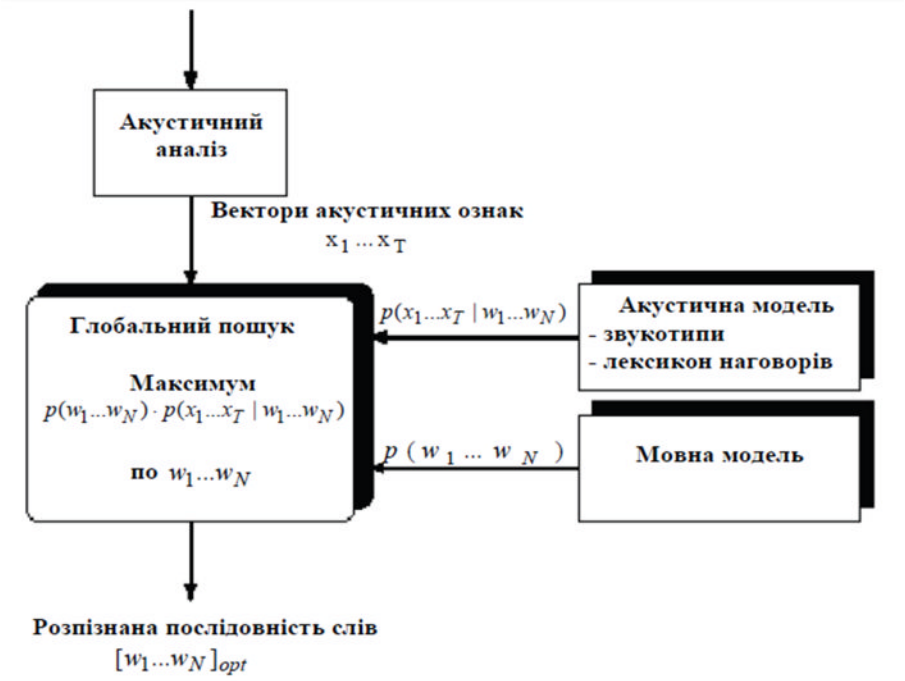

Рис. 1.3. Архітектура системи автоматичного розпізнавання мови

Архітектура системи ґрунтується на структурі правил прийняття рішень Байєса та включає параметризацію мовного сигналу, визначення ймовірностей акустичних та мовних характеристик та глобальний пошук для визначення послідовності слів з найвищою ймовірністю. Такий метод мінімізує кількість правок для перетворення вимовленого в текст, спираючись на природний та загальноприйнятий критерій рішення.

Методика даного методу описана наступним чином, а саме: позначимо

$$W = \{w_1, w_2, ..., w_n \} \tag{1.3}$$

де W є множиною, що містить N слів, і $O = \{o_1, o_2, ..., o_T \}$ є набором векторів спостереження, на основі яких пристрій приймає рішення щодо розпізнавання вимовлених слів. Якщо $P(W/O)$ є ймовірністю того, що при спостереженні ознак W O вимовлені слова W, то пристрій повинен прийняти рішення на користь W за умови

$$P(\hat{W}/O) = \max_{W} P(W/O). \tag{1.4}$$

Процес перетворення мовного сигналу в дані O називається акустичною обробкою. В системах АРМ для обробки мовної інформації використовуються непараметричні та параметричні методи.

РОЗДІЛ 2
НАУКОВО-ДОСЛІДНА ЧАСТИНА

2.1 Математичні моделі процесів формування та сприйняття мови

Точна математична модель мовоутворення описує артикуляцію, не лише розширює наше загальне розуміння процесу мовоутворення людиною, але й створює можливості для імітації цього процесу в машинному синтезі мови.

Дослідження особливостей мовного сигналу виявляє його складний характер, що ускладнює створення точних моделей мовоутворення. З цього приводу існуючі моделі базуються на спрощеному розумінні процесу формування мови.

При моделюванні синтезу мовного сигналу враховуються ключові компоненти артикуляційної системи. У даному розділі проводиться аналіз відомих математичних моделей мовоутворення з метою виокремлення їхніх основних переваг та недоліків.

2.2 Математичні моделі утворення мови

Однією з найбільш розповсюджених математичних моделей мовоутворення є лінійна модель [2], представлена на рис. 2.1. Відповідно до цієї моделі - звуковий тиск в області біля губ при відомому тиску на виході джерела звукових коливань визначається за допомогою відповідного рівняння:

$$Y(\omega) = P_{(j}(\omega)V_{(j}(\omega)V_S(\omega)V_{l,}(\omega), \qquad (2.1)$$

де ω – є значенням частоти коливань, а $V_{(i}(\omega), V_s(\omega), V_i(\omega)$ - це значення передачі для джерела звуку, мовного тракту і випромінювача відповідно.

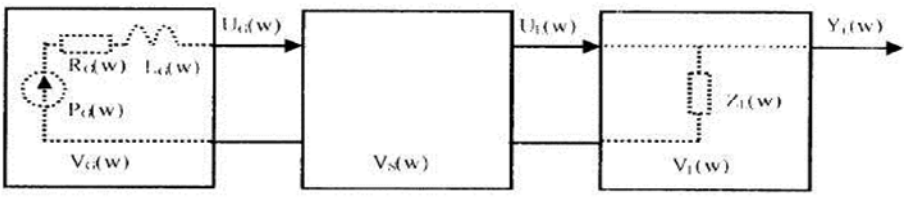

Рис. 2.1. Лінійна модель мовоутворення

Значення передачі джерела, мовного тракту і випромінювача визначаються:

$$G_x(\omega) = \frac{1}{Z_{(i}(\omega_{(} = \frac{1}{R_{(i} + j\omega L_{(i};$$

$$Z_L(\omega) = \frac{j\omega L_L R_L}{R_L + j\omega L_L} \quad (2.2)$$

де R_L і ωL_L - відомі значення активного і індуктивного опорів голосової щілини, а також активного індуктивного опорів випромінювача. Передаточна функція мовного тракту $V_S(\omega)$ визначає його резонансні властивості і обчислюється числовим інтегруванням хвильових рівнянь для визначеної конфігурації мовного тракту.

Згідно з правилом (2.2), ключовими ознаками для визначення мовних звуків є форма інстантного (миттєвого) спектру сигналу та характеристики формантів з їх часовими властивостями. У реальних умовах всі ці параметри визначаються за допомогою аналізатора у формі смугових фільтрів, які здійснюють таке перетворення:

$$F_S(j\omega, t) = \int_{-\infty}^{\infty} r_S(\tau - t) y(\tau) e^{-j\omega\tau} d\tau \quad (2.3)$$

де $F_S(j\omega,t)$ - інстантний (миттєвий) спектр сигналу, $r_S(t)$ - функція ковзної ваги часу [4].

Друга широко використана модель у галузі мовоутворення - модель лінійного прогнозу, що є аналогічна лінійній моделі в частині залежності від z-області [2]:

$$Y_L(z)) = P_{(j}(z) V_M(z), \quad (2.4)$$

де Z є комплексною експонентою, $Y_L(z), P_G(z) i V_M(z)$ - z є значенням перетворення мовного сигналу, сигналу збурення і передаточного значення системи, що визначають $V_M(z)$ загальний спектр та враховують вплив випромінювання, мовного тракту і збурення:

$$V_M(z) = -\frac{G_S}{1 - \sum_{K=1}^{M'} a_k z^{-k}}, \quad (2.5)$$

де G_S і M' - коефіцієнти передаточних функцій підсилення та прогнозування, та визначають порядки відповідних чисельника та знаменника. Ця модель була розроблена для ефективного виділення інформативних ознак мовних виразів безпосередньо в області часу за допомогою обчислювальних пристроїв. Основна концепція методу лінійного прогнозу полягає в тому, що кожен наступний дискретний відлік мовного сигналу може бути наближено лінійною комбінацією попередніх відліків:

$$y^*(i) = \sum_{k=1}^{M'} a_k y(i-k). \tag{2.6}$$

де a_k - невідомі коефіцієнти, що вираховуються шляхом мінімізації середньоквадратичного відхилення між фактичними значеннями сигналу та прогнозованими [3].

Слід зауважити, що аналіз за допомогою "лінійного прогнозу" вважається одним із найадекватніших методів аналізу голосового сигналу через його точність та ефективність обчислення. Головна концепція кодування за допомогою лінійного прогнозу (ЛПК) полягає в тому, що зразок голосового сигналу може бути наближено лінійною комбінацією попередніх зразків. Шляхом мінімізації квадратичної різниці між реальним зразком голосового сигналу та передбаченим сигналом можна визначити коефіцієнт прогнозу, який суттєво впливає на лінійну комбінацію цих попередніх зразків.

В процесі реалізації лінійного прогнозу виникає низка проблем через необхідність:

- визначення кількості використовуваних коефіцієнтів прогнозу.
- ідеальності інтервалу для аналізу.
- локалізації інтервалу, що аналізується, без порушення періоду основної частоти.
- врахування ефекту кінцевої довжини робочих слів у системі попередньої обробки.

2.3 Математичні моделі сприйняття мови

Механізм слухового сприйняття залишається ще недостатньо вивченим в цілому. Сучасні знання фізіології вуха, електрофізіології нервових клітин та суб'єктивної поведінки аудитора у психоакустичних випробуваннях дозволяють прослідкувати взаємозв'язок між певними функціями слуху та цими різними галузями знань. Завдяки кількісній оцінці та аналітичній передбачуваності поведінки, вдалося покращити розуміння механізмів слухового сприйняття.

Перший крок у цьому напрямку включає побудову математичної моделі, яка описує зсув базилярної мембрани під впливом звукового тиску біля барабанної перетинки. На верхньому зображенні на рис. 2.2 наведено спрощену схему периферійних органів слуху, яка лягла в основу математичного моделювання.

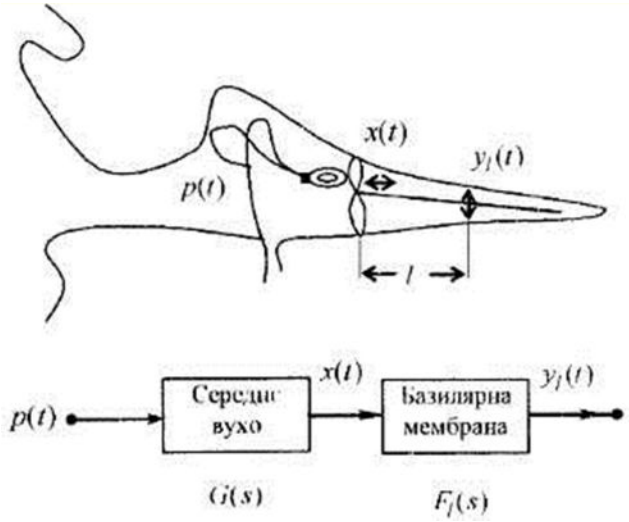

Рис. 2.2 Схематичне зображення вуха

На цій спрощеній схемі вушний равлик зображено в розгорнутому вигляді: $p(t)$ - тиск звуку біля барабанної перетинки, $x(t)$ - еквівалентний лінійний зсув основи стремінця, та $y_l(t)$ - лінійний зсув базилярної мембрани в точці, розташованій на відстані l від стремінця.

Основною метою при розробці моделі мовосприйняття є визначення приблизної аналітичної залежності між вказаними параметрами. Зручність вирішення цієї задачі забезпечується її поділом на два етапи. Початково апроксимується передаточна функція середнього вуха, що визначає зв'язок між $x(t)$

та $p(t)$. На другому етапі проводиться апроксимація передаточної функції системи на ділянці від стремінця до вказаної точки l на мембрані, що встановлює залежність між $x(t)$ і $p(t)$. Функції апроксимації представлені на рис. 2.2 у вигляді частотних перетворень $G(s)$ і $F_l(s)$ відповідно. Вибір функцій $G(s)$ і $F_l(s)$ повинен відповідати фізіологічним даним. Якщо вважати, що механічна система вуха у цільовому діапазоні частот є пасивною і лінійною, для апроксимації фізіологічних даних можна використовувати раціональні функції частоти з стабільними спектральними максимумами (полюси в лівій півплощині). Окрім зручності розрахунків, раціональні функції володіють властивостями фільтра низьких частот. Враховуючи, що модель встановлює взаємозв'язок між вхідним та вихідним сигналами, описуючи передавальні властивості системи, можна знехтувати докладними розрахунками для інших точок при розрахунках реакції в обраній точці мембрани.

2.4 Результативність систем розпізнавання мови: аналіз ефективності на лексичному та синтаксичному рівнях: огляд та аналіз

Для оцінки інформативності систем розпізнавання мови корисним є використання Байєсівського класифікатора. У цілях цього використання будуються вирішальні функції, представлені наступним чином:

$$d_k(W^*) = P(W^*) \cdot P(W_k \mid W^*), \tag{2.7}$$

де: d_k - вирішальна функція приналежності образу до класу $k - го$:

$P(W^*)$ - ймовірність спостереження мовного повідомлення W^* за умови його належності до класу;

$P(W_k \mid W^*)$ - ймовірність приналежності мовного образу W^* до класу W_k.

При розробці СРМ важливим є припущення про існування таких комбінацій параметрів, що дозволяють мінімізувати помилки розпізнавання в системі загалом. Для урахування цього в теорії складних систем часто використовується принцип мінімізації ентропії, яка тісно пов'язана з рівнем помилок розпізнавання [26].

Розглядаючи інформаційний канал, в який надходять мовні сигнали W, а на виході отримуються повідомлення W^*, задачу побудови потенційної СРМ можна сформулювати наступним чином [4]:

Дано: Канал передачі інформації мовного повідомлення, що складається з наступних даних:

- словник слів української мови - W_d;
- ймовірності подібності у словнику слів - $p(w_i / w_j)$;
- рівень шуму навколишнього середовища (дБ) - r_d.

Завдання: Визначити кількість інформації, яку потенційна система розпізнавання мови отримує.

Основне поняття в даній моделі становить інформація, яка спостерігається на виході дискретного або неперервного випадкового процесу [9]:

$$l(W_k) = \log\frac{1}{P(W_k)} = -\log P(W_k),\ \ l(W^*) = \log\frac{1}{P(W^*)} = -\log P(W^*). \tag{2.8}$$

Кількість інформації, яку отримує потенційна система розпізнавання визнається:

$$\mathrm{I}_{\Pi} = \mathrm{H}_{\Pi}(W) - \mathrm{H}_{\Pi}(W/W^*), \tag{2.9}$$

де $\mathrm{H}_{\Pi}(W)$ - ентропія вхідного повідомлення до його передавання, пов'язана з апріорною ймовірністю окремого повідомлення $P(W_i)$ таким виразом:

$$\mathrm{H}_{\Pi}(W) = \sum_{i=1}^{|W|} P(w_i) \cdot l(w_i) \cdot \log_2 P(wi) \tag{2.10}$$

Апріорна ймовірність $P(w_i)$ визначається шляхом статистичного аналізу або за умови рівної ймовірності для всіх повідомлень, у такому випадку:

$$P(w_i) = \frac{1}{|W|}. \tag{2.11}$$

де $|W|$ - кількість повідомлень у словнику W: $\mathrm{H}_{\Pi}(W/W^*)$. Ентропія повідомлення після його прийому вказує на втрату інформації для одного повідомлення, визначається як:

$$\mathrm{H}_{\Pi}(W/W^{*}) = -\int\sum_{i=1}^{|W|}P(w_i/w^{*})\log P(w_i/w^{*})dx = \sum_{w_i=W}\sum_{w_i=W} p(w_i/w_j)p(w_j)\log_2\frac{p(w_i)p(w_i/w_i)}{\sum_{w_i}p(w_i/w_j)p(w_i)} \qquad (2.12)$$

Величини $p(w_i/w_j)$, що вказують на подібність слів, можуть бути експериментально визначені для різних рівнів завад на основі результатів розрізнення слів і фонем людиною.

Імовірність сплутування $p(w_i/w_j)$ між словами w_i і w_j обчислюється за допомогою формули:

$$p(w_i/w_j) = \frac{\tilde{\mu}_i \cdot 10^{-d_{ij}^{w}}}{\sum_{k=1}^{|w|}\tilde{\mu}_k \cdot 10^{-d_{ij}^{w}}}, \qquad (2.13)$$

де d_{ij}^{w}- відстань між *i*-им і *j*-тим словами словника, а $\tilde{\mu}_k$ - середнє значення відстані для слова w_j.

Відстань d_{ij}^{w} можна розрахувати, якщо відома матриця відстаней $\| d_{ql}^{m_a} \|$ між основними елементами розпізнавання (фонемами або звукотипами):

$$d_{ij}^{W} = \sum_{k=1}^{m_w} d_{i_k j_k}^{m_a}(r_d), \qquad (2.14)$$

де $d_{i_k j_k}^{m_a}(r_d)$ - відстань між $k-го$ фонемою чи звукотипом $i-го$ слова і $k-ю$ фонемою чи звукотипом $j-го$ слова при заданому рівні шумів r_d; m_w - кількість фонем у слові.

З виразу (2.14) випливає висновок про те, що моделлю СРМ може бути матриця відстаней $\| d_{ql}^{m_a}$ між фонемами, а також оператор перетворення $\phi(r)$, який здійснює визначення рівня впливу шуму на цю матрицю.

Відстань між словами словника можна також обчислити за допомогою методів динамічного програмування. Відстань d_{ij}^{w} між двома словами, які мають різну кількість фонем $T1$ і $T2$ *(T1=T2)*, розраховується як сума локальних відстаней $d_{qJ} = d(i_q \cdot j_l)$ вздовж шляху вирівнювання між послідовностями фонем; локальна функція відстані $d(\cdot\cdot)$ реалізується за допомогою евклідової метрики. Для

обчислення кумулятивної відстані між двома словами використовується рекурсивна формула:

$$D_{l_q j_l} = \begin{cases} \min\left\{D_{i_{q-1}, j_{l-1}}, D^{0,}_{\infty}{}_{i_{q-1}, j_l} D_{i_q, j_{l-1}}\right\} + d^{m_a}_{i_q j_l} & \text{якщо } i = j = 0; \\ & \text{якщо } i>0, j>0 \\ & \text{в іншому випадку} \end{cases} \quad (2.15)$$

Окремі функції відстаней $d^{W}_{ij} = D_{i_{T1} j_{T2}}$ обчислюються за стандартними обчислювальними операціями $T1 \cdot T2$. Дані для побудови такої моделі отримуються з матриць подібності (сплутування) фонем $|\delta_{ij}|$, які визначаються на основі слухових експериментів з сприйняття фонем людиною.

У матриці подібностей (i, j)-ий елемент представляє собою ступінь сприйняття мовного стимулу S_i як стимулу S_j, де S_j- різні фонеми з фонемної множини $S^{m_a} = \{S^{m_a}_{l}, ... S^{m_a}_{m_n}\}$.

Основою для створення потенційної системи розпізнавання фонем є метод багатовимірного неметричного масштабування [5], який дозволяє відновити метричну конфігурацію простору фонем з неметричних величин подібності δ_{ij}. Задачею багатовимірного масштабування є знаходження відображення:

$$F : S^{m_a}_{N_0} \to V_T \mid (\delta_{ikjK} \geq \delta_{iKjK}) \Rightarrow (d_{iljl} \leq d_{sKjK}), \quad (2.16)$$

де обмеження називається умовою монотонності; d_{iKjK}- відстань між точками ϑ_{iK} і ϑ_{jK} у просторі $V_T = \{\vartheta_1, ... \vartheta_2\}$, що береться з частково упорядкованої множини відстаней $D = \{d_{i1j1}, ... d_{iMjM}\}$; N_0- розмірність простору-прообразу; $M_0 = N_0(N_0 - 1)/2$ - кількість пар точок у просторі. Ідеальним вважається такий метричний простір, відстань в якому узгоджується з заданими розходженнями якнайкраще (під розходженням розуміється зворотний ранговий порядок величин подібності $(\delta_{шо})$). Ступінь монотонності відношення між відстанями і розходженнями можна визначити за допомогою різних функцій нормалізованих сум квадратів помилок:

$$Q = \frac{\sum_{i \le j} (d_{ij} - \bar{d}_{ij})^2}{\sum_{i<j} \bar{d}_{ij}^2}, \quad (2.17)$$

де d_{ij} - усереднене значення відстані.

Оскільки Q є функцією розташування точок ϑ_i у просторі, теоретично проблема багатовимірного масштабування зводиться до мінімізації функції багатьох змінних $Q(\vartheta_1, \vartheta_2, ..., \vartheta_N)$ (в дійсності, критерій узгодження Q є функцією $(N_o \cdot T)$ змінних, оскільки кожний вектор ϑ_1 у просторі-образі має T координат). Мінімізація функції багатьох змінних виконується за допомогою стандартної процедури градієнтного спуску. Алгоритм чисельного методу градієнтного спуску, який реалізує метод багатовимірного неметричного масштабування, обирає довільну конфігурацію точок ϑ_k у просторі V_T і потім здійснює ітеративний процес зміни векторів ϑ_k у напрямку градієнта функції критерію:

$$grad Q(\bar{\vartheta}_K) = \frac{2a_0}{\sum_{i<j} \bar{d}_{ij}^{-2}} \sum_{j \ne K} (d_{Kj} - \bar{d}_{Kj}) \frac{\bar{\vartheta}_K - \bar{\vartheta}_j}{d_{Kj}}, \quad (2.18)$$

де a_0 - константа; $(\bar{\vartheta}_K - \bar{\vartheta} j) / d_{Kj} = \frac{(\bar{\vartheta}_K - \bar{\vartheta}_j)}{|\bar{\vartheta}_K - \bar{\vartheta}_j|}$ - градієнт вектора у напрямку від ϑ_K до ϑ_j: $|x|$ - норма величини x.

Градієнт (2.18) обчислюється на кожному кроці ітерації та застосовується для оновлення координат векторів.

Для оцінки впливу шуму на якість роботи потенційної системи, експериментальні дані обиралися для шести рівнів співвідношень "сигнал/шум" - 5; 0; 5; 10; 15; 20 дБ. Зашумлення оригінального мовного сигналу виконувалося шляхом накладення білого шуму в програмному середовищі Matlab. У ролі мовного матеріалу використовувались звуки української мови, які зачитувалися п'ятьма дикторами, а сприйняття записаного мовного матеріалу здійснювалось десятьма особами.

Обробка отриманих матриць подібностей фонем за різних відношень "сигнал/шум" за допомогою алгоритму багатовимірного неметричного масштабування привела до отримання матриць відстаней між фонемами $|d_{ij}^{m_a}|$ для значень r = {-5, 0, 5,10,15, 20} дБ. Отримані результати свідчать, що вплив шуму на простір звукотипів можна математично описати лінійним оператором перетворення:

$$\| d_{ij}^{m_a} \|_r = \Phi^{m_a}(r) \| d_{ij}^{m_a} \|_{r_0} , \quad (2.19)$$

де $\| d_{ij}^{m_a} \|_r$ - матриця відстаней для відношення "сигнал/шум" r_0 = 0 дБ.

Аналогічно оператор $\Phi(r)$ може бути обчислений як:

$$\Phi^h(r) = C_0 + \sum_i C_i \cdot r^i , \quad (2.20)$$

де C_i - константи, що мають свої значення для всякого алфавіту елементів мови, що класифікуються; r - задане відношення "сигнал/шум", дБ. Константа залежна від r [9].

У ході експериментальних досліджень була проведена апроксимація для різних типів звуків української мови, таких як голосні, приголосні, дзвінкі приголосні, глухі приголосні, сонорні приголосні. Функція апроксимації мала вигляд степеневого поліному з порядковим номером від 1 до 3.

У таблиці 2.1 подані формули для обчислення оператора $\Phi(r)$ для голосних і приголосних звуків, а також для визначення значення точності апроксимації R^2. Визначено, що для моделювання потенційної системи розпізнавання мови голосні звуки можуть бути описані лінійним рівнянням, приголосні звуки при рівні шуму r, такого що $r > 5$, можуть бути описані рівнянням другого порядку, а для рівня шуму $r < 5$, - рівнянням третього порядку.

Таблиця 2.1 Аналітичний вигляд оператора $\Phi(r)$ для голосних та приголосних звуків.

Ступінь функції	Голосін звуки	
	Рівняння	R^2
1	0,0028x+0,9355	0.90
2	$--10^{-4}x^2+0{,}0043x+0{,}932$	0.96
3	$-10^{-7}x^3+10^{-4}x^20{,}0039x+0{,}932$	0.98
Ступінь функції	Приголосні звуки	
	Рівняння	R^2
1	0,0144x+0,07753	0.74
2	$-0{,}0009x^{-2}+0{,}0246x+0{,}79$	0.98
3	$-4*10^{-5}x^3-0{,}0017x^20{,}026x+0{,}81$	0.99

На основі експериментів і даних з таблиці 2.1 були побудовані графіки залежності достовірності розпізнавання звуків української мови. Рисунок 2.3 відображає усереднені значення відстаней між фонемами в залежності від рівня значення"сигнал/шум".

З використанням отриманих відстаней між фонемами можна обчислити відстані між словами заданого словника розпізнавання (за формулою 2.13) і визначити інформативність системи на акустичному рівні для заданого співвідношення "сигнал/шум".

Граматика представляє собою поверхневу структуру мовного повідомлення і включає фонетичну структуру слів та порядок слів у реченні. Для побудови граматичної структури автоматичної СРМ застосовується теорія формальних мов. Формальна мова відрізняється від спонтанної людської мови визначеним набором символів та правил виведення з цього набору речень. Однією з ключових особливостей спілкування з машиною є те, що користувач може створювати лише ті повідомлення, які відомі системі (обмежена кількість слів або речень).

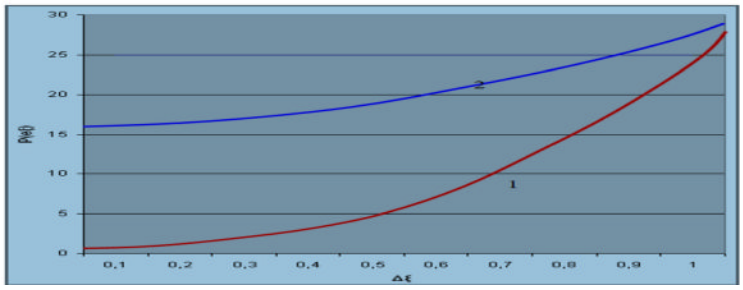

Рис. 2.3. Взаємозв'язки між точністю розпізнавання звуків і рівнем шуму в навколишньому середовищі.

Розглянемо мову $L(G)$, яка породжується граматикою G. При розробці автоматизованої системи розпізнавання мови обирається словник термінальних символів граматики (звукотипів, фонем, складів, слів або словосполучень) в залежності від кінцевого призначення системи. Ефективність обраної термінальної бази для системи розпізнавання можна оцінити за допомогою інформаційних характеристик обраної формальної мови [7].

Ентропія мови $H(L(G))$ може бути точно визначена з діаграми переходів між станами системи. Ключовою кількісною величиною, що описує діаграму, є її матриця зв'язності, яку визначають за виразом:

$$C_{ij=z},\ i,j=i,j=1,\vec{N}\ , \tag{2.21}$$

z - кількість переходів зі стану q_i в стан q_j.

Матриця зв'язності використовується для визначення кількості речень N_k довжиною k у мові $L(G)$ за допомогою виразу:

$$N_k = \vec{e}_1 C^k \vec{\int}\ , \tag{2.22}$$

де $\vec{e}_1$ представляє перший N - компонент одиничного вектора, $\vec{\int} N$ - компонентний вектор із одиниць або нулів, i - та компонента якого дорівнює 1, якщо q_i є кінцевим станом. Оскільки

$$\sum_{k=()}^{\infty} C^k = (l-C)^{-1}\ , \tag{2.23}$$

де l - одинична матриця розміром N x N, то загальна кількість речень у мові $|L(G)|$ дорівнює

$$|L(G)| = \vec{e}_l (l - C)^{-1} \int^{\rightarrow T} \tag{2.24}$$

і середня довжина речення

$$|\overline{W}| = \frac{\sum_k kN_k}{|L(G)|}. \tag{2.25}$$

При умові $|L(G)| < \infty$ і рівноймовірності речень можна розрахувати ентропію мови $N_k, |L(G)|, |\overline{W}|$ за формулою:

$$H(L(G)) = \frac{\log_2 |L(G)|}{|\overline{W}|}. \tag{2.26}$$

Якщо умова рівноймовірності речень не виконується, то максимальну ентропію можна отримати при будь-якому ймовірнісному розподіленні речень:

$$H_{\max}(L(G)) = -\log_2(x_0), \tag{2.27}$$

де x_0 обрано з того розрахунку, щоб $1 - \sum_k N_k x_0^{\,k} = 0$.

Після визначення ентропії мови можна розглядати ентропію системи розпізнавання в цілому. Для цього СРМ можна уявити як канал передачі інформації (рис. 2.4). Основною характеристикою цього каналу є втрата інформації на одне слово $H(x/y)$, що є мірою невизначеності реалізації слова x при його розпізнаванні як слова y (згідно з (2.2)). Поняття $H(x/y)$ може бути розширено і на речення за допомогою виразу:

$$H(W/\overline{W}) = |\overline{W}| H(x/y), \tag{2.28}$$

втрата інформації на одне речення рівнозначна добутку ентропії на одне слово на середню кількість слів в реченні. Остаточно, з нерівності Фано отримуємо:

$$H(W/\overline{W})\eta \le \frac{H(P_e)}{|\overline{W}|} + P_e H(L(G)), \tag{2.29}$$

$$H(P_e) = P_e \log_2 P_e + (1 - P_e)\log_2(1 - P_e),$$

де P_e - ймовірність помилки розпізнавання речення, а η - ефективність мови

$$\eta = \frac{H(L(G))}{\log_2(|V_T|)}, \eta \le 1. \tag{2.30}$$

Рівняння (2.29) при відомих параметрах $H(W/W)$, $\eta.|W|$. $H(L(G))$ дозволяє знайти P_e. Оскільки (2.29) є трансцендентним рівнянням, він розв'язується чисельними методами. Приведемо рівняння (2.29) до вигляду для вирішення його ітераційним методом [15]:

$$\begin{matrix} H(W/W)\cdot & \eta\_P_{e\cdot}\cdot\log_2 P_e + (1-P_e)\cdot\log_2(1-P_e)_{\le P_e} \\ H(L(G)) & |W|\cdot H(L(G)) \end{matrix} \tag{2.31}$$

Очевидно, що область рішень цього рівняння знаходиться в інтервалі [0;0,5][0;0,5].

Приведемо, наприклад, розрахунок передбачуваної помилки розпізнавання для словника, що складається з 10 цифр англійської мови. Для експерименту визначення матриці сплутування між цими словами залучено дев'ять дикторів, які вимовляли кожне слово по 10 разів. Половина отриманої вибірки слів була використана для навчання дев'яти відповідних марківських мереж лінійного типу з трьома емісійними станами, а інша половина була використана для визначення ймовірностей сплутування слів цього словника.

Для опису граматики необхідно визначити операції, які можна вирішувати з її терміналами. Розглянемо два випадки:

1. Граматика дискурсу складається з десяти речень довжиною в один термінал.

Grammar "G_1"

< main(result) >:< digit(result) >;

< digit(result) >: one("result=1")|

two("result=2")|

nine("result=9")|

zero("result=0")|

2. Граматика складається зі 100 речень довжиною в два термінали.

Grammar "G_2"

< main(result1, result2) >:< digit(result) > < digit(result2)

< digit(result) >: one("result=1")|

zero("result=0")|

На рис. 2.4 наведені графічні результати визначення ймовірності помилки розпізнавання для обох випадків.

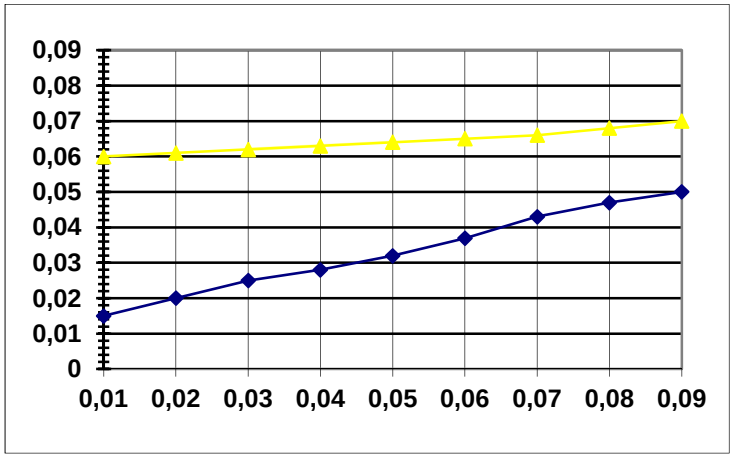

Рис. 2.4. Графічний спосіб визначення помилки розпізнавання

Після прийняття повідомлення інформаційним каналом для першого випадку ентропія становить 2,66 біт/речення (3,32 - до прийняття повідомлення), а для другого випадку - 5,32 біт/речення (6,64 до прийняття повідомлення). Відповідно, інформація, яку отримала така потенційна система розпізнавання, складає 0,66 і 1,32 біт відповідно.

2.5. Огляд математичної моделі призначеної для оптимізації процесу класифікації голосових команд

Реалізація процесу класифікації голосових команд для обширного словника використовує багатомодульні системи розпізнавання. Кожен модуль спрямований на виявлення та/або розпізнавання конкретного рівня звукотипів мови. Структура модулів визначається на підставі експертної інформації. На сьогоднішній момент не існує математичної моделі для оптимального вибору конфігурації модульної системи розпізнавання мови, що враховує умови конкретної задачі.

Формалізація мети оптимізації процесу класифікації може бути виражена у вигляді:

$$\tilde{S}_{Gopt} = \arg\max Э(\tilde{S}_{Gi}) \quad \left\{ \tilde{S}_{Gi} \in \tilde{S}_{G,} W_{d,} r_{d,} E_d \right\} \tag{2.32}$$

де стратегія $\overline{S}_{Gi}$ - одна з стратегій розпізнавання зі здійсненого множини, дані $\tilde{S}_{G,} W_{d,} r_{d,} E_d$ - інформація, доступна на час класифікації (словник, рівень завад, тощо), а умови задачі включають в себе параметри, такі як словник, рівень завад і точність класифікації.

Складність вирішення поставленої задачі оптимізації, вираженої у вигляді (2.32), визначається тим, що обчислювальна складність C_p критерію ефективності (2.7) виступає функцією двох змінних - витрат на процедуру класифікації C_k та витрат на обчислення ознакового опису C_x. Розкладання цих змінних у критеріях (2.7) можна виконати, представивши стратегію розпізнавання у вигляді послідовної процедури класифікації.

Розглянемо будь-який модуль у ієрархії процесу класифікації.

Основними елементами *i*-го функціонального блоку є:

- S^i - алфавіт, що описує голосові команди для розпізнавання.
- $\bar{l}(S^i)$ - параметрична характеристика мовного сигналу (математична модель на етапі попередньої обробки), визначається як:

$$\bar{l}(S^i) = x^i_{opt}, x_{opt} = \arg\min C_i(x^i_l); \tag{2.33}$$

- $C_i(x^i_l)$ - обчислювальна складність *1-ї* ознаки для опису мовних станів.
- $A_i(\bar{l}(S^i))$ - математична модель прийняття рішення (класифікатор).
- $\tau^i cp$ - середній час класифікації.

Ці компоненти ілюструють кроки при здійсненні процесу класифікації голосових команд, які можна розглядати окремо для подальшої декомпозиції та оптимізації.

Процес розпізнавання полягає в знаходженні оптимальної послідовності кроків класифікації:

$$S_{Gopt} = A_1(\tilde{l}(S^1)) \otimes A_2(\tilde{l}(S^2)) \otimes ... \otimes A_W(\tilde{l}(S^W) \tag{2.34}$$

Процес розпізнавання мови може бути відображений у вигляді дерева класифікації, де існує один кореневий канал, а всі інші канали є термінальними. Кількість гілок у такому дереві (етапів класифікації) відповідає кількості термінальних каналів, які представляють розпізнавані образи.

Для розкриття принципів створення оптимального дерева класифікації необхідно визначити його характеристики. Для цього розглянемо приклад довільного дерева рішень, представленого на рис. 2.5.

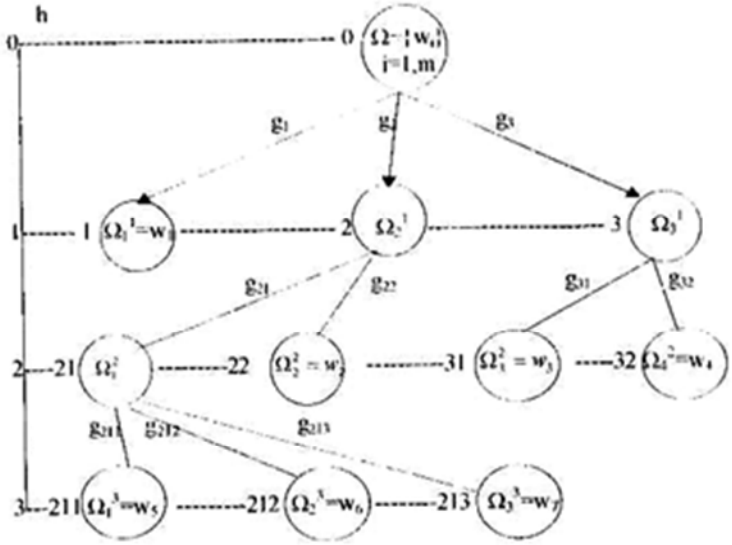

Рис.2.5.Приклад дерева класифікації

На зображеному дереві $T_r = (\Omega, G)$ міститься множина вузлів $\Omega = \{\Omega_i^h\}$ (класів образів, на які здійснено розбиття множини образів на даному рівні дерева) та множина гілок $G = \{g_{ki}\}$, де h - висота дерева, i - номер класу образів на висоті h, k - індекс вузла-попередника *j-го* нащадка. Індекс k формується рекурсивною конкатенацією індексів всіх вузлів-попередників з метою створення можливості однозначної ідентифікації всього шляху рішень, який передував розпізнаванню образа, зображеного термінальним символом дерева.

Наприклад, для розпізнавання образу w_5, зображеного термінальним вузлом з індексом 211, необхідно прийняти рішення про класифікацію в вузлах з індексами 0, 2, 21. Дерево класифікації на рис. 1.8 включає сім термінальних вузлів $w_1, w_2, ..., w_7$ з індексами 1, 22, 31, 32, 211, 212, 213 відповідно, три внутрішні вузли $\Omega_2^1, \Omega_3^1, \Omega_1^2$ з індексами 2, 3 і 21 відповідно, і один кореневий канал Ω з індексом 0. Кореневий вузол відповідає словнику, який розпізнається.

Позначимо так, щоб кожному довільному дереву класифікації відповідали:

C_A - кількість різних гілок дерева, що пропорційна часу класифікації і відображає обчислювальну складність алгоритму класифікації;

t_A - середній час класифікації образу деревом рішень;

$P(w_i)$ - апріорна ймовірність образу w_i;

T_C - кодове дерево, ізоморфне дереву рішень T_r.

Додатково введемо числові параметри для кодового дерева T_C:

η_i - число висячих вершин кодового дерева, що відповідає числу образів M, що розпізнаються;

m_c - кількість ребер в кодовому дереві, що дорівнює кількості гілок дерева;

$m_C = C_A$ - середня довжина кодового дерева, що відповідає середньому часу класифікації t_A;

P_η - ймовірність висячих вершин кодового дерева, що дорівнює апріорним ймовірностям $P(w_i)$ термінальних вузлів дерева рішень з однаковими індексами;

Для деякого внутрішнього вузла Ω_i^h дерева, що має k класів, з якого виходить m вузлів-нащадків, зберігається умова адитивності ентропії:

$$H_K = H_m(P_1, P_2, ..., P_m) + \sum_{i=1}^{m} p_i * H_{ni}(q_{i1}, q_{i2}, ..., q_{in_i}), \tag{2.35}$$

де n_i - кількість синів *i-го* вузла-нащадка; p_i- ймовірність цього вузла, $\sum_{i=1}^{m} P_i = 1$;

q_{ij} ймовірність *j-го* сина *i-го* вузла-нащадка, $\sum_{j=1}^{n_i} q_{ij} = 1$, *i=1, m.*

Це рівняння описує, як ентропія внутрішнього вузла дерева є сумою ймовірностей його нащадків, помножених на їхню власну ентропію.

Формулу (2.35) можна легко вивести, розглядаючи ентропію *i-го* вузла-нащадка. Ймовірності p_i його синів дорівнюють добутку ймовірності складної події, яка складається з двох $P_i \cdot q_{ij}$ незалежних подій. Тоді ентропія, що відповідає *i-й* гілці термінального вузла, дорівнює

$$H_l = P_i q_{i1} \log_2 P_i \sum_{j=1}^{n_i} q_{ij} + P_i \sum_{j=1}^{n_i} q_{ij} \log_2 (P_i q_{i2}) + ... + P_i q_{in}, \log 2 (P_i q_{in_l}) =$$

$$= P_i \log_2 P_i \sum_{j=1}^{n_i} q_{ij} + P_i \sum_{j=1}^{n_i} q_{ij} \log_2 q_{ij} = H(P_i) + P_i H(q_{ij}) \qquad (2.36)$$

де *k* - індекс вузла-нащадка, що вказує на конкретний клас образу в даному термінальному вузлі, а p_i- ймовірність появи конкретного класу вузла-нащадка відносно *i-го* вузла.

Враховуючи, що термінальний вузол має *m* вузлів-нащадків, і підсумовуючи ентропію по кожному *i-му* вузлу, можна отримати (2.35).

На основі цього висновку має місце така теорема:

Теорема 1.1. Якщо на будь-якому рівні *h* дерева T_r виконуються умови

$$\Omega_i^h \bigcap \Omega_j^h = \varnothing, \forall i \neq j,$$

$$H^{(h)} = -\sum_{i=1}^{m_h} P_i^{(h)} \cdot \log_2 P_i^{(h)} \geq H_0 \qquad (2.37)$$

для кожного внутрішнього вузла *h-го* рівня, то середній час t_A класифікації невідомого образу від кореневого вузла до термінального задовольняє наступному співвідношенню

$$t_a \leq H_w / H_0 \qquad (2.38)$$

У (2.39) і (2.40) використані такі позначення:

H_0 - порогове значення ентропії для вузлів дерева рішень;

$P_i^{(h)}$ - ймовірність, пов'язана з вузлом Ω_i^h, $P_i^h = \sum_{j \subset \Omega_i^h} P_j$;

$H_w = -\sum_{i=1}^{M} P(w_i) \cdot \log_2 P(w_i)$ - ентропія множини образів, що розпізнаються;

$P(w_i)$ - апріорна ймовірність класу w_i: m_h кількість гілок вузла Ω_i^h на рівні *h*.

Теорему 2.1 доведено за допомогою математичної індукції, враховуючи, що кожен внутрішній вузол дерева разом із своїми нащадками також є деревом рішень (піддеревом).

Вирішення завдання вибору оптимального коефіцієнта розгалуження суттєво обмежує обсяг пошуку в процедурі оптимізації дерева рішень. Припустимо, що середня висота термінальних вузлів дорівнює *h*, а середній коефіцієнт розгалуження B_r. Тоді кількість різних шляхів *M* від кореня дерева до термінальних вузлів

$$M = B_r^h, \tag{2.39}$$

звідки можна визначити

$$B_r^h =^h j_M. \tag{2.40}$$

З виразу H_0 можна визначити, що значення наближено дорівнює

$$H_0 \approx \log_2(B_r). \tag{2.41}$$

Оскільки кількість вузлів, які виходять з певного внутрішнього вузла, наближено дорівнює B_r, то з формули (2.41) можна записати середній час класифікації образу в дереві як

$$t_A = B_r^{\bar{h}} \cdot \frac{H_w}{\log_2 B_r}. \tag{2.42}$$

Мінімізація виразу (1.47) показує t_A, що ВВ мінімальне при B_r=2.7182…

Теорема 2.2. Якщо помилка e^h класифікації образу в кожному вузлі дерева на рівні *h* задовольняє умову $e^h \le e_0$, де e_0 - деяке граничне значення помилки, то загальна помилка класифікації E_d образу в дереві визначається виразом:

$$E_d = e_0 \cdot \frac{H_w}{H_0} \tag{2.43}$$

Доведення: Верхня границя помилки класифікації в дереві визначається як $E_0 = e_0 \cdot t_A$, звідки, використовуючи (2.40), маємо $E_d \le E_0 = e_0 \frac{H_w}{H_0}$, що і потрібно довести.

Оскільки $H_w \approx \log_2 M$, а $H_0 = \log_2 B_r$, то

$$E_d \le e_0 \cdot \frac{\log_2 M}{\log_2 B_r}. \tag{2.44}$$

Одночасна мінімізація як помилки класифікації (2.44), так і часу класифікації (2.40) визначає обмеження коефіцієнта розгалуження B_r, який вибирається при побудові оптимального дерева рішень:

$$2 \le B_r \le 5 \tag{2.45}$$

Отже, знайдені характеристики дерева рішень дозволяють обмежити область пошуку при розв’язанні задачі оптимізації (2.28).

Задачу розробки моделі для оптимізації процесу розпізнавання голосових команд можна вирішити шляхом застосування оптимізаційного методу "керованого пошуку вперед з поверненням" [6]. У цьому методі критерій (2.1) визначає напрямок пошуку структури дерева рішень серед всіх можливих конфігурацій. На кожному етапі пошуку обирається така конфігурація каналів, яка має найвище значення критерію. Для заданого вузла Ω_i^h дерева процедура пошуку має такий вигляд:

1. Одне з можливих розбиттів $\pi^h \in \Pi$ вузла Ω_i^h на підмножину нащадків $\{\Omega_j\}, j = 1, m.$ виконується на основі вибраної ознаки $x^h \in X$. Ознака x^h обирається за матрицею відмінності таким чином, щоб коефіцієнт розгалуження лежав в межах, визначених у (2.45).

2. Отримана конфігурація вузла піддається обчисленню значення критерію (2.2).

3. Проводячи повторні кроки 1 і 2, формуються інші можливі розбиття, а для кожного з них визначається середнє значення критерію.

4. Визначається конфігурація, при якій значення критерію має максимальне значення. Таким чином, встановлюється оптимальний набір ознак для даного каналу дерева та оптимальний крок при алгоритмі класифікації.

РОЗДІЛ 3
ТЕХНОЛОГІЧНА ЧАСТИНА

3.1 Використання прихованих марківських моделей (ПММ) у системах розпізнавання мови

Застосування статистичних методів у сфері розпізнавання образів визначає важливий етап у розвитку автоматичного розпізнавання мови (АРМ). Це призвело до використання розгалуженої математичної статистики та теорії ймовірностей, що істотно підвищило точність процесу розпізнавання. На сучасному етапі практично всі визнані СРМ ґрунтуються на статистичних методах.

В рамках цього підходу мовний сигнал розглядається як випадковий образ, який слід розпізнати, або, точніше сказати, перетворити у послідовність слів *W*. Як результат, завдання розпізнавання мовних сигналів формулюється як класична задача класифікації знаків з використанням критерію максимальної апостеріорної ймовірності. З іншого боку, це означає максимізацію апостеріорної ймовірності, визначеної як $P(W \mid X)$, де X - це ряд акустичних векторів, які описують параметри мовного сигналу, а W - послідовність слів. Згідно з формулою Байєса, апостеріорну ймовірність можна зобразити у вигляді:

$$\arg\max_{w \in \Gamma} P(W \mid X) = \arg\max_{w \in \Gamma} P(X \mid W) \cdot P(W) \qquad (3.1)$$

де Γ - множина всіх можливих послідовностей слів, $P(W \mid X)$ - умовна ймовірність появи послідовності акустичних векторів X для заданої послідовності слів W, $P(W)$ - апріорна ймовірність виникнення конкретної послідовності акустичних векторів W. $P(W \mid X)$, - вираз зазвичай використовується для позначення акустико-фонетичної моделі, а $P(W)$ - моделі мови [8].

На сьогоднішній день широкого поширення серед технологій акустико-фонетичного моделювання мовного сигналу набрало використання прихованих марківських моделей (ПММ). Використання ПММ надає можливості досягти

високої точності розпізнавання, забезпечує ефективне відтворення мовного сигналу і становить потужний та гнучкий інструмент для створення СРМ. Незважаючи на ці переваги, ПММ мають свій набір обмежень, таких як обмежена дискримінантна потужність, особливо при використанні критерію максимуму правдоподібності (МП). В інших випадках, наприклад, при використанні критерію максимуму взаємної інформації (МВІ), можна досягти більшої роздільної здатності, але це призводить до математично складніших алгоритмів та вимагає значних обмежень. Додатково, використання акустичної та фонетичної контекстуальної інформації призводить до ускладнення самої структури ПММ, великих обсягів пам'яті для зберігання параметрів моделі і значної кількості навчальних даних.

Практично всі СРМ, розроблені протягом останніх двадцяти п'яти років, базуються на статистичних принципах використовують ПММ. Основні концепції даної теорії були сформульовані і опубліковані на зламі 60-х і 70-х років XX століття у серії статей Баума та інших вчених [12]. Перші практичні результати використання ПММ в системах автоматизованого розпізнавання мови були представлені Бейкером [9] та Єлінеком з колегами в IBM [2]. Згодом було написано кілька оглядових статей, які полегшили використання теорії ПММ у практичних застосуваннях [4, 5].

Для прикладу розглянемо модель для звуку, яка базується на марківській моделі, зображеній на рис. 3.1. Ця модель включає в себе послідовність станів, позначених як $S_1, S_2, ..., S_5$, і пов'язаних миттєвими ймовірнісними переходами, зображеними стрілками з ймовірністю переходу між станами. Переходи можуть відбуватися лише між певними станами, при цьому можливе зациклення (зокрема *i-го* стану *j-с)*. Модель здійснює ймовірнісний перехід з одного стану в інший або залишається в тому ж стані, видаючи вихідний акустичний вектор y_k з вихідним ймовірнісним розподілом $b_n(y_k)$, що відповідає цьому стану. Ці ймовірності відомі під назвою емісійні ймовірності. Таким чином, вислів, який описує послідовність акустичних векторів параметрів $X = \{x_1, x_2, ..., x_n\}$, може бути промодельований як послідовність дискретних стаціонарних станів із миттєвими переходами між ними

$Q = \{q_1, q_2, ..., q_K\}, K < N$, та послідовністю згенерованих акустичних векторів $Y = \{y_1, y_2, ..., y_N\}$.

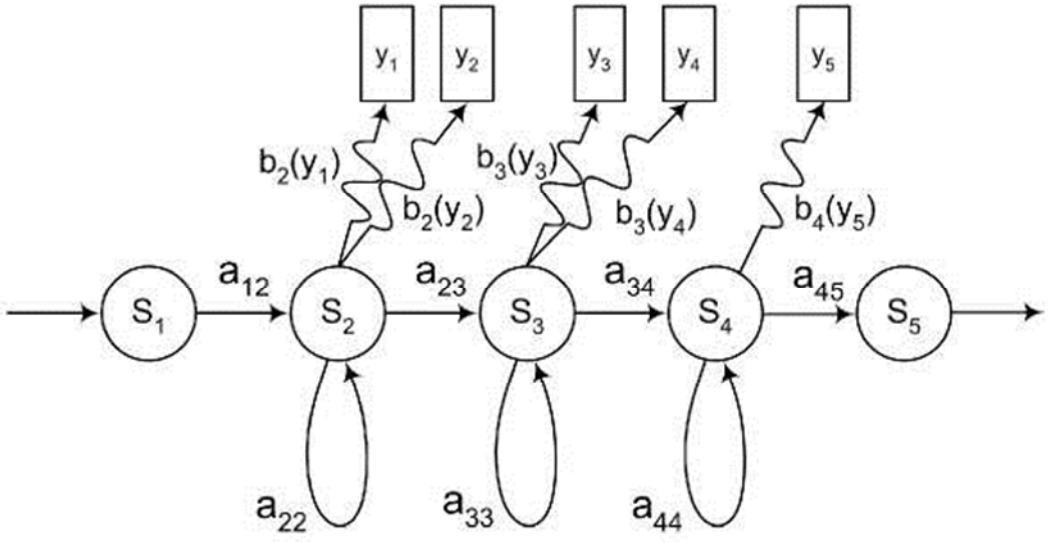

Рис. 3.1. Зображення системи ПММ

Отже, ПММ складається з марківського ланцюга зі скінченною кількістю станів S_N і матриці перехідних (транзитивних) ймовірностей a_{ij}, яка визначає тривалість стану системи перебування в кожному конкретному стані. Іншими словами, марківський ланцюг моделює тимчасові зміни мовного сигналу. Крім того, модель включає скінченну множину емісійних ймовірностей $b_n(y_k)$ для моделювання спектральних варіацій сигналу. Цей підхід передбачає наявність двох одночасних стохастичних процесів. Перший - головний і неспостережний (прихований) - це послідовністю станів ПММ. Другий - це випадковий процес, що визначається послідовністю спостережень. Ця модель називається "прихованою" марківською моделлю, оскільки ми можемо спостерігати лише випадковий процес, а не сам марківський ланцюг [1].

Для опису ПММ потрібно задати наступні компоненти:

1. Множина станів моделі $S = \{S_1, S_2, ..., S_N\}$ (де N - є кількістю станів моделі). Стан моделі в момент часу t позначається q_t.

2. Множину різних символів спостереження, які можуть бути породжені моделлю $Y = \{y_1, y_2, ..., y_K\}$ (де K - кількість символів спостереження моделі). Символи спостереження відповідають фізичному виходу модельованої системи.

3. Матриця ймовірностей переходів між станами $A = \{a_{ij}\}$, де

$$a_{ij} = P[q_{t+1} = s_j \mid q_t = s_i] \quad 1 \le i, j \le N, \tag{3.2}$$

Тут a і j - ймовірності переходів зі стану i в стан j, що залежать від часу.

4. Множина розподілів ймовірностей появи символів спостереження (ці ймовірності називають емісійними або вихідними) в стані $j, B = \{b_j(k)\}$ де $b_j(k) = P[y_k$ в момент $t[q_t = s_j], 1 \le j \le N, 1 \le k \le K$.

5. Початковий розподіл ймовірностей початкових станів $\Pi = \{\pi_i\}$

$$\pi_i = P[q_1 = s_i], \quad 1 \le i \le N \tag{3.3}$$

Для застосування прихованих марківських моделей в системі АРМ необхідно зробити кілька спрощень, які є необхідним припущеннями про мовний сигнал.

- Послідовні спостереження є статистично незалежними - це означає, що ймовірність послідовності спостережень є просто добутком ймовірностей окремих спостережень.
- Хоча мовний сигнал є нестаціонарним процесом, його можна описати як послідовність векторів спостережень, які є кусково-стаціонарними процесами.
- Ймовірність перебування в конкретному стані в момент часу *t* залежить лише від стану, в якому знаходився процес в момент часу *t-1*.

Тепер розглянемо просту СРМ. Ідеально використовувати приховані марківські моделі для кожного вірогідного висловлювання. Проте, очевидно, що це є практично неможливим для широкого спектру завдань, наприклад розпізнавання ізольованих команд із великого мовного словника. Для цього використовують менші одиниці мови, такі як фони, які лінгвістично є відповідними фонемам. Для кожної одиниці мови необхідно створити окрему ПММ, де $M = \{m_1, m_2, ..., m_U\}$ - множина ПММ для всіх можливих фонів, а $\Theta = \{\lambda_1, \lambda_2, ..., \lambda_U\}$ - множина пов'язаних з ними

параметрів. Отже, M_i представляє марківську модель одного слова, яку отримано конкатенацією елементарних моделей із множини M, де M_i складається з L_i які є станами $q_l \in S$ і $l = 1,2,...,L_i$, а множиною параметрів цієї моделі є Λ_i, підмножиною Θ.

Опис кожного фону визначається за допомогою послідовності векторів спектральних ознак сигналу. Під час навчання для кожного слова (наприклад M_i) створюється послідовність, яка включає множину повторень векторів параметрів X_{Mi}, що відповідають вимові цього слова декількома дикторами. Нашою метою є вибір множини параметрів Θ, яка максимізує ймовірність $P(M_i \mid X_{M_i}, \Theta)$ для всіх навчальних висловлювань X_{Mi}, що стосуються M_i, тобто

$$\arg\max \prod_{i=1}^{l} P(M_i \mid X_{M_i}, \Theta) \tag{3.4}$$

Отже, процес навчання передбачає оптимізацію параметрів моделі на основі конкретного критерію ефективності. На сьогоднішній день, відсутній відомий аналітичний вираз для визначення цих параметрів. Додатково, на практиці, коли у наявності є доступна послідовність спостережень у вигляді навчальних даних, відсутній чіткий метод для оптимального їх оцінювання. Проте, за допомогою ітеративних процедур, таких як метод математичного очікування-максимізації (EM), алогоритм Баума-Уелча чи градієнтних методів можна налаштувати параметри моделі так, щоб локально максимізувати ймовірність $P(M \mid X_M, \Theta^*)$. Важливо відзначити, що ці алгоритми відносяться до категорії алгоритмів "без учителя" при процесі навчання, оскільки вони забезпечують оцінку параметру розподілу ймовірностей без необхідності попередньої розмітки. Під час етапу розпізнавання невідомого виразу необхідно вибрати найбільш підходящу модель M_i яка максимізує ймовірність $P(M \mid X_M, \Theta)$ при фіксованих параметрах Θ та заданій послідовності спостережень X. Отже, результатом розпізнавання вислову X буде слово, яке відповідає моделі M_i, що максимізує відповідну ймовірність:

$$i = \underset{\forall j}{\arg a \max}\, P(M_j \mid X, \Theta \tag{3.5}$$

Метод пошуку оптимальної моделі, відомий як алгоритм Вітербі, бере за основу процедуру динамічного програмування. Процес навчання і розпізнавання мовних сигналів пов'язані із здійсненням вибору конкретного критерію оптимальності. Всі вони мають конкретний фізичний зміст і знаходять застосування на практичній діяльності. Вибір оптимальності здійснюється з урахуванням таких критеріїв як максимум апостеріорної ймовірності чи найбільшої правдоподібності, що визначають параметри окремої моделі, обсяг тренувальних даних і обчислювальних витрат та впливає на результат точності розпізнавання та здатності до узагальнення даних з тренувальної вибірки. Одним із найефективніших критеріїв є Байєсівський класифікатор, що ґрунтується на апостеріорній ймовірності $P(M_i \mid X_{M_i}, \Theta)$ (рідше максимальній апостеріорній ймовірності, MAP-оцінювач). Цей класифікатор враховує той факт, що вибір послідовності акустичних векторів був генерований моделлю M_i з великим вибором параметрів Θ. З використанням даного правила Байєса $P(M_i \mid X_{M_i})$ цей підхід можна виразити наступним чином:

$$P(M_i \mid X, \Theta) = \frac{P(X \mid (M_i \Theta) P(M_i \mid \Theta)}{P(X \mid \Theta)}, \tag{3.6}$$

яке здійснює розділення завдання процесу оцінки ймовірності на дві складові: акустичного та мовного $P(M_i \mid \Theta)$ моделювання.

Мовне моделювання має на меті оцінювати апріорні ймовірності висловлювань $P(M_i \mid \Theta)$. Ця модель мови зазвичай вважається незалежною від акустичних моделей і описується за допомого множини набору параметрів Θ. Параметри моделювання зазвичай оцінюються на обширних текстових базах даних [9].

$$\frac{P(X \mid (M_i \Theta)}{P(X \mid \Theta)} \tag{3.7}$$

Акустичне моделювання має своїм завданням оцінку щільності ймовірностей, що здійснюється незалежно від інших моделей. Оскільки ймовірність $P(X \mid M_i \Theta)$ обумовлена лише M_i, вона залежить тільки від параметрів моделі M_i і упускається

$P(X \mid \Theta)$, вираз (3.7) можна переписати так: $P(X \mid (M_i \Lambda_i)$, де Λ_i - множина параметрів, пов'язаних з M_i моделлю. Отже, як частину процесу навчання і розпізнавання, необхідно оцінити ймовірність $P(X \mid (M_i \Lambda_i)$. Ця ймовірність відома як глобальна правдоподібність послідовності векторів параметрів X при заданій M_i.

Дані $P(X \mid (M_i \Lambda_i)$ отримуємо як результат додавання

$$P(X \mid (M_i \Lambda_i) = \sum_{\{\Gamma_i\}} P(X, \Gamma_i \mid M_i, \Lambda_i), \tag{3.8}$$

в якому поняття $\{\Gamma_i\}$ охоплює усі можливі шляхи, або послідовності станів, довжини яких дорівнює L в моделі M_i. Для кожної такої послідовності станів ймовірність отримання послідовності спостережень $X_1^L = \{x_1, x_2, ..., x_L\}$ визначається рівнянням

$$P(X_1^L \mid q_1^L, M_i \Lambda_i) = \prod_{l=1}^{L} P(x_l \mid q_1^l, X_1^{l-1}, M_i, \Lambda_i) \tag{3.9}$$

Можна зробити висновок, що розрахувати вираз (3.9) можна за допомогою алгоритму прямого-зворотнього ходу. Для цього потрібно рекурсивно обчислити пряму змінну [17]

$$P(q_l^n, X_1^n \mid M_i, \Lambda_i) = \sum_{k=1}^{L} P(q_k^{n-1}, X_1^{n-1} \mid M_i, \Lambda_i) p(q_1^n, x_n \mid q_k^{n-1}, X_1^{n-1}, M_i, \Lambda_i), \tag{3.10}$$

де $(q_l^n, X_1^n \mid M_i, \Lambda_i)$ являє собою ймовірність того, що підпослідовність спостережень $X_1^n = \{x_1, x_2, ..., x_n\}$ частково створила модель M_i, а в період часу n спостерігався стан $q_l^n = S_l$ та був визначений вектор спостережень x_n.

Другий добуток у правій частині рівняння (3.10) можна виразити як добуток ймовірностей

$$P(q_l^n, X_1^n \mid q_k^{n-1}, X_1^{n-1} \mid M_i, \Lambda_i) = p(x_n \mid q_l^n, q_k^{n-1} \mid M_i, \Lambda_i) p(q_l^n \mid, q_k^{n-1}, M_i, \Lambda_i), \tag{3.11}$$

Перший множник $p(x_n \mid q_l^n, q_k^{n-1} \mid M_i, \Lambda_i)$ визначає ймовірність емісії, а другий $p(x_n \mid q_l^n, q_k^{n-1} \mid M_i, \Lambda_i)$ - ймовірність транзиції між станами. Зазвичай для спрощення

емісійну ймовірність зводять до вигляду $p(x_n \mid q_l)$, припускаючи, що спостережуваний акустичний вектор залежить тільки від поточного стану .

Описана типова модель прихованої марківської моделі є ефективним інструментом, який дозволяє розробникам значно покращити якість розпізнавання мовного сигналу. Це підтверджується численними лабораторними дослідження систем розпізнавання змішаної мови з великими обсягами словників (від 1000 до 40000 слів), які показують високі результати в порівняльних випробуваннях, проведених у рамках проекту SQALE.

Завершуючи короткий опис ПММ, слід відзначити їхні незаперечні переваги:

- Потужний математичний апарат: Використання складних математичних концепцій і методів.
- Ефективне моделювання тимчасових та спектральних варіацій мовного сигналу: Здатність точно відтворювати як часові зміни, так і спектральні характеристики мови.
- Гнучка топологія: Можливість легкого включення фонологічних та синтаксичних правил, а також конструювання моделей слів з основи, використовуючи моделі фонів.
- Глибоке практичне опрацювання: Розроблення потужних навчальних та розпізнавальних алгоритмів, які забезпечують ефективне навчання на великих мовних базах даних та розпізнавання ізольованих слів і злитого мовлення без адаптації під конкретного диктора.

Дослідження ПММ виявили низку недоліків:

- Слабкі дискримінантні здібності: В процесі навчання акустичні моделі формуються на основі критерію максимуму правдоподібності, що призводить до обмеженої точності, оскільки не використовується більш точний критерій максимуму апостеріорної ймовірності.
- Статистична незалежність векторів спостережень: Мовний сигнал, навпаки, характеризується кореляцією в часі, але ПММ припускають статистичну незалежність векторів спостережень.

- Кусково-постійний характер моделі: Кожен марківський стан має стаціонарну статистику, що не завжди відповідає динаміці мовлення змінної тривалості.
- Апріорний вибір топології та розподілів: Неможливість ефективно визначити оптимальну топологію та розподіли заздалегідь.
- Відсутність адекватних моделей тривалості станів: Відсутність ефективних і адекватних моделей тривалості станів, які враховують природу мовного сигналу.
- Марківська модель першого порядку: Обмеженість, оскільки стан в момент часу *n* обумовлений лише попереднім станом системи в момент часу *n-1*.
- Навчання та оптимізація лінгвістичної моделі відбувається окремо від акустичної: Недостатня інтеграція навчання лінгвістичних та акустичних моделей.

Ці недоліки суттєво обмежують потенціал цього класу моделей та підштовхують дослідників до пошуку альтернативних та доповнюючих підходів для вирішення проблем акустико-фонетичного моделювання мовного сигналу.

3.2 Використання штучних нейронних мереж (ШНМ) у системах розпізнавання мови

Інший клас моделей, які використовуються для акустико-фонетичного моделювання, - це моделі штучних нейронних мереж (ШНМ). З середини 1980-х років ШНМ почали широко застосовуватися в СРМ [9]. Головною перевагою, яка зумовила широке використання ШНМ, є їх потужні дискримінантні властивості, можливість навчання та представлення неявних знань. Однак, незважаючи на їх потенційні можливості у класифікації короткочасних фонетично-акустичних одиниць (фонем), ШНМ не стали основною моделлю для систем АРМ. Причиною цьому був недолік ШНМ, пов'язаний зі складністю моделювання тривалих послідовностей спостережень, таких як слова і висловлювання, через їхню суттєву тимчасову варіативність. Цю проблему не вдалося вирішити навіть застосування рекурентних архітектур мережі. Іншими словами, ШНМ ефективно функціонують

лише зі статичними образами, і їхня ефективність значно падає, коли на вході присутня динаміка, така як нелінійні зміни у часі.

Структура та принципи роботи ШНМ базуються на біологічних моделях нервових систем, зокрема, на концепціях головного мозку. Розглядати нейронні мережі можна як форму самоорганізації, представляючи собою набір однотипних та паралельно функціонуючих елементів, нейронів, які спілкуються один з одним та взаємодіють з зовнішнім середовищем через спеціально організовані зв'язки. В моменти часу інформація передається від вхідних зв'язків до нейрону, на основі якої формується вихідний сигнал згідно з певними принципами. Цей вихідний сигнал подається на входи інших нейронів або взаємодіє з "зовнішнім світом".

Основним будівельним блоком нейронних мереж є нейрон. Модель нейрона МакКаллока-Пітса (рис. 3.2), запропонована у 1943 році [13], є широко використовуваною. Відповідно до даної моделі нейрон має певний набір вхідних зв'язків і один вихід, який може бути розпаралелений. Роботу цієї моделі можна описати наступним чином: до входу нейрона подається вхідний вектор $x(t)=\{x_1(t),x_2(t),...,x_N(t)\}$, який скалярно множиться на вектор ваг $w_k=\{\omega_{1k},\omega_{2k},...,x_{Nk}\}$. Іншими словами, компоненти вектора $x_i(t)$ зважуються ваговими коефіцієнтами ω_{ik} згідно з формулою:

$$u_k(t)=\sum_{i=0}^{N}\omega_{ik}x_i(t) \qquad (3.12)$$

.

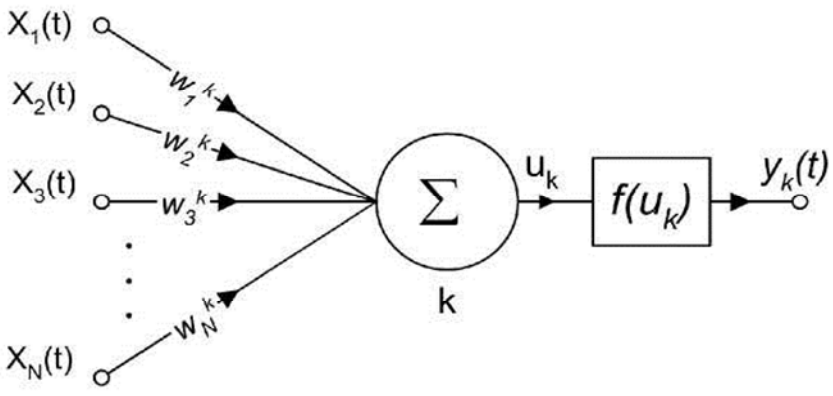

Рис. 3.2. Модель нейрона

Вихідний сигнал до нейрона можна визначаити як

$$u_k(t) = f(u_k(t)), \tag{3.13}$$

де $f(u_k(t))$ це функція активації нейрона. Найчастіше в якості функції активації вибирається нелінійна безперервна функція, зокрема сигмоїдальна функція

$$f(x) = \frac{1}{1+e^{-ax}}. \tag{3.14}$$

де параметром, що впливає на форму функції активації і обирається користувачем, є *a*.

Найвідомішою та широко використовуваною моделлю нейронної мережі є багатошаровий перцептрон (БП), структурна схема якого представлена на рис. 3.3.

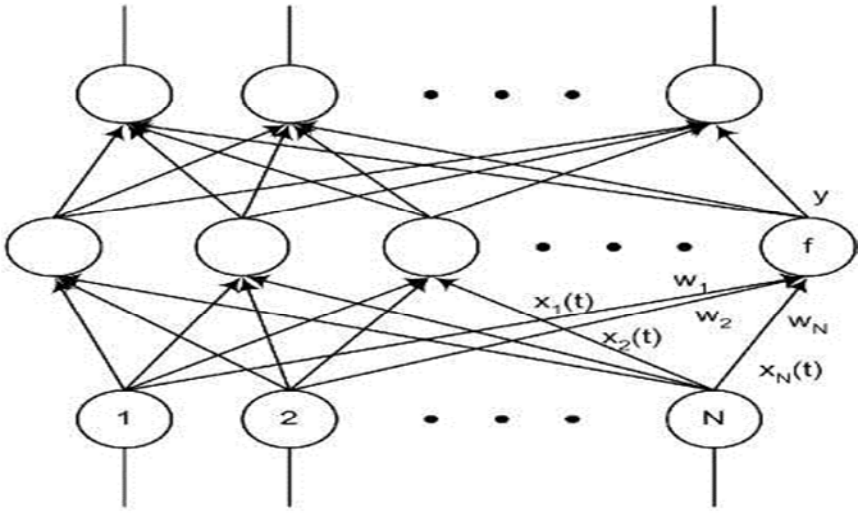

Рис. 3.4. Тришаровий персептрон

Елементи багатошарового персептрона розділені на кілька шарів, всередині кожного шару елементи є лінійно впорядковані і невзаємодіючі між собою. Кожен нейрон, що знаходиться в мережі (за винятком нейронів у вхідному прошарку, які є рецепторами) отримує вхідний сигнал від кожного нейрона попереднього шару, а вихідний (крім останнього шару) надходить на вхід до нейронів наступного шару. Таким чином, багатошаровий персептрон є моделлю, в якій зв'язки забезпечують передачу сигналу тільки в одному напрямку від входу до виходу, без зворотного шляху.

Елементи проміжних шарів отримали назву "приховані елементи", а їхні шари - "приховані шари". Самі нейрони, як правило, функціонують відповідно до моделі МакКаллока-Пітса, а сигмоїдальна функція (3.14) часто обирається в якості функції активації. Найбільш відомим алгоритмом навчання багатошарового персептрона є процедура, яку описав Розенблатт у 1959 році, модифікація цього алгоритму, відома як алгоритм зворотного поширення помилки (Back Propagation Error). Цей алгоритм дозволяє здійснювати контрольване навчання (навчання "з вчителем").

Алгоритм зворотного поширення помилки (ВР-алгоритм) представляє собою градієнтний метод оптимізації, спрямований на мінімізацію функції вартості (цільової функції) між бажаним та згенерованим виходом мережі. Основною метою навчання є встановлення бажаного функціонального відношення між даними вхідними та вихідними шляхом коригування вагових коефіцієнтів між нейронами.

В процесі навчання, вибравши певні початкові значення ваг, вхідний та бажаний вихідний (цільовий) вектори одночасно подаються на вхід мережі. Мережа виконує вхідне відображення вектора на вихідний. Різниця між отриманим та цільовим векторами визначає помилку ε_k:

$$\varepsilon_k(t) = u_k^{trg}(t) - f(w_k(t), x(t)), \tag{3.15}$$

де $u_k^{trg}(t)$ - цільовий вихід *k-го* нейрона на *t-м* кроці алгоритму, $w_k = \{\omega_{1k}, \omega_{2k}, ..., x_{Nk}\}$ - ваговий коефіцієнт між *k-им* і *j-им* нейронами, *x(t)* - вхідний вектор, *f()* - нелінійна функція активації нейрона ε_k. Для підтримки корекції ваг використовується градієнт $\{\omega_{kj}\}$ цільової функції від виходу до входу мережі.

Алгоритм ВР може використовувати різні цільові функції, зокрема, середньоквадратичну похибку.

$$E = \sum_{t=1}^{T} \| y(t) - y^{trg}(t) \|^2, \tag{3.16}$$

або функцію відносної ентропії

$$E_e = \sum_{t=1}^{T} \sum_{k=1}^{K} \left[u_k^{trg}(t) \ln \frac{u_k^{trg}(t)}{u_k} + (1 + u_k^{trg}(t)) \ln \left(\frac{1 - u_k^{trg}(t)}{1 - u_k(t)} \right) \right], \tag{3.17}$$

У вищенаведеному виразі, $u_k^{trg}(t)$ представляє цільовий вихід *k-го* нейрона вихідного слою на *k-м* кроці алгоритму, $u_k(t)$ - спостережуваний вихід *k-го* нейрона вихідного слою, *K* - нейронна кількість у вихідному шарі, і *T* - загальна кількість навчальних образів.

Важливим етапом при здійсненні процесу навчання мережі є метод корекції зв'язків ваг. Оскільки навчання виконується за допомогою методу найшвидшого спуску, коригування ваг зв'язків проводиться в напрямку від'ємного градієнта цільової функції. У рівнянні (3.18) використовується символ η, який представляє крок навчання, і відповідає швидкості збіжності алгоритму найшвидшого спуску.

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \eta \frac{\vartheta E}{\vartheta \omega_{ij}(t)} x_i(t), \tag{3.18}$$

Слід відзначити, що використання градієнтних методів для навчання багатошарового персептрона (БП) може гарантувати досягнення лише локального мінімуму на поверхні цільової функції, що може виявитися значно віддаленим від глобального мінімуму. Виходження з області локального мінімуму застосуванням простого алгоритму найшвидшого спуску є неможливим. Для подолання цієї проблеми часто використовують метод з моментом.

У методі з моментом процес модифікації ваг визначається не лише інформацією про градієнт функції, але й фактичним трендом змін ваг $\Delta\omega_{ij}$, що обчислюється за допомогою коефіцієнта моменту:

$$\Delta\omega_{ij}(t+1) = -\eta \frac{\vartheta E}{\vartheta \omega_{ij}(t)} x_i(t) + a\Delta\omega_{ij}(t), \tag{3.19}$$

де перший доданок відповідає звичайному методу модифікації ваг, а другий - моменту, який відображає останню зміну ваг і не залежить від фактичного значення градієнта. Коефіцієнт моменту *a* вибирається в інтервалі $0 < a < 1$. Зауважте, що вплив моменту особливо помітний неподалік від локального мінімуму, де значення градієнта зближується до нуля, що сприяє виходу з області локального мінімуму. Проте великий вплив моменту (при великих значеннях) може викликати нестабільність, тобто розбіжність алгоритму навчання.

У перших експериментах одношаровий персептрон продемонстрував вражаючі результати при навчанні в простих нелінійних завданнях. Можна сказати, що одношаровий персептрон, як класифікатор образів, формує в просторі ознак дискримінантні гіперплощини. Ці гіперплощини, при перетині класів образів і за умови слабкої нелінійної порогової функції, мінімізують середньоквадратичну помилку між фактичними u_k і очікуваними u_k^{trg} виходами [5].

Іншими словами, у випадку двох класів, образи яких розподілені за нормальним законом і за умови, що ознаки, які описують образи, є некорельованими, можна побудувати одношаровий персептрон з такою ж вирішальною функцією, як і в параметричному гауссовому класифікаторі.

Але слід відзначити, що одношаровий персептрон не здатен розділити образи, які потребують більш складних поверхонь в просторі ознак. Наприклад, він не може вирішити проблему виключення АБО, побудувавши просту гіперплощину.

Зі збільшенням кількості шарів класифікаційні властивості персептрона якісно покращуються. Двошаровий персептрон може вирішити проблему виключення АБО формуючи опуклу розділяючу поверхню (що є результатом перетину гіперплощин, утворених елементами першого шару), але його можливості обмежені. Двошаровий персептрон не може успішно представляти або апроксимувати функції поза дуже вузьким і специфічним класом.

Використання тришарового персептрона відкриває ще більше можливостей для апроксимації відображення з одного простору у інший. Тришаровий персептрон може формувати розділяючі поверхні більш складної форми і отримувати будь-які наперед задані безперервні функції вхідних сигналів. Зокрема, вибравши відповідну вирішальну функцію, він може емулювати будь-який традиційний детермінований класифікатор [7].

Теоретичні основи для цих висновків надає результат А. Н. Колмогорова про можливість представлення будь-якої дійсної неперервної функції змінних у вигляді суперпозиції скінченного числа безперервних дійсних функцій з глибиною вкладення не більше трьох, що використовують лише лінійне підсумовування аргументів і безперервно зростаючі функції однієї змінної. Це дозволяє

багатошаровим персептронам ефективно моделювати складні взаємозалежності між вхідними та вихідними даними.

Основними стимулюючими факторами для використання багатошарового персептрона є наступні переваги нейронних мереж:

- Персептрон може проводити дискримінантне навчання між мовними одиничними елементами, що представляють вихідні класи персептрона. Навчання персептрона включає оптимізацію параметрів для кожного класу на відповідних даних, а також відхилення від даних, що належать іншим класам.
- Персептрон може знаходити оптимальну комбінацію обмежень для класифікації, при цьому немає потреби у строгих припущеннях про розподіл вхідних ознак, що часто вимагається в стандартних методах математичного програмування.
- Персептрон є структурою з високим рівнем паралелізму, що сприяє використанню паралельного обладнання.

3.3 Огляд нейронних мереж із затримкою часу (TDNN)

TDNN мережа представляє собою спробу використовувати статичний багатошаровий персептрон для розпізнавання динамічних тимчасових послідовностей мовних даних. Це можна досягнути шляхом перетворення часової послідовності в просторову послідовність відповідних нейронів. У TDNN кожному моменті часу на нейрони вхідного шару подається не лише поточний вектор параметрів *x(t)* в момент часу *t*, але й частина послідовності векторів з запізненням $X_{t-k}^{t} = \{x(t-k), x(t-(k-1)),...,x(t)\}$ та випередженням $X_{t}^{t+k} = \{x(t), x(t+1)),...,x(t+k)\}$. Таким чином, активність кожного нейрона прихованого шару залежить від активності вхідного шару нейронів на певному кінцевому часовому інтервалі X_{t}^{t+k}. Аналогічно вихідний шар зв'язаний з прихованим шаром. Як видно з рис. 3.3, активність вихідного нейрона визначається активністю нейронів з скритого шару, взятих в момент часу *t1, t, t+1*. Кількість кроків, на які TDNN "заглядає" вперед і назад в

часі, визначається розробником моделі. Для процесу навчання мережі з такою топологією також може використовуватися алгоритм зворотного поширення помилки (BP).

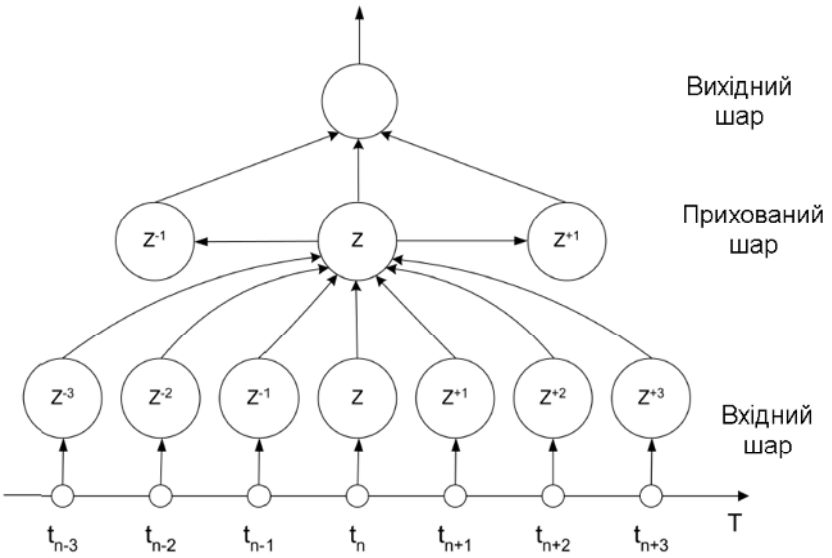

Рис. 3.3. Нейрона мережа з затримкою часу

Вайбел був одним із перших, хто досліджував цю модель. Ленг та Хінтон провели експеримент з використання TDNN для розпізнавання ізольованих звуків "B, D, E, V" без адаптації під конкретного диктора. Для навчання мережі використовувався акустичний матеріал, зібраний від 100 чоловічих дикторів. У результаті досягнута точність 7.8% помилок. Подальші експерименти з синтезом модульної мережі, в якій кожен окремий модуль представляв собою TDNN-мережу, спеціалізовану на розпізнаванні звуків, показали можливість надійної ідентифікації всіх прийнятних звуків для японської мови ізольовано вимовлених дикторами - японцями. Точність розпізнавання в цих експериментах досягла 95.9%. При цьому точність розпізнавання голосних звуків у тих же експериментах досягла 98.6%.

3.4 Огляд рекурентних мереж (RNN)

Ще одним способом використання контекстуальної інформації є встановлення взаємозв’язків між нейронами незалежно від їх топології в мережі. Однак, щоб зберегти властивості звичайної зворотної передачі, ці зв'язки повинні мати затримку

на один часовий крок. Такі зв'язки називають рекурентними. Структуру такої мережі представлено на рисунку 3.4. Відповідно до даної моделі, активність нейронів залежить від активності нейронів на попередньому рівні та від активності нейронів попереднього часового кроку. Такий тип мережі відомий як рекурентна або динамічна нейронна мережа (RNN). Початково RNN застосовулась в СРМ з обмеженими успіхами через труднощі з навчанням, аналізом та розробкою.

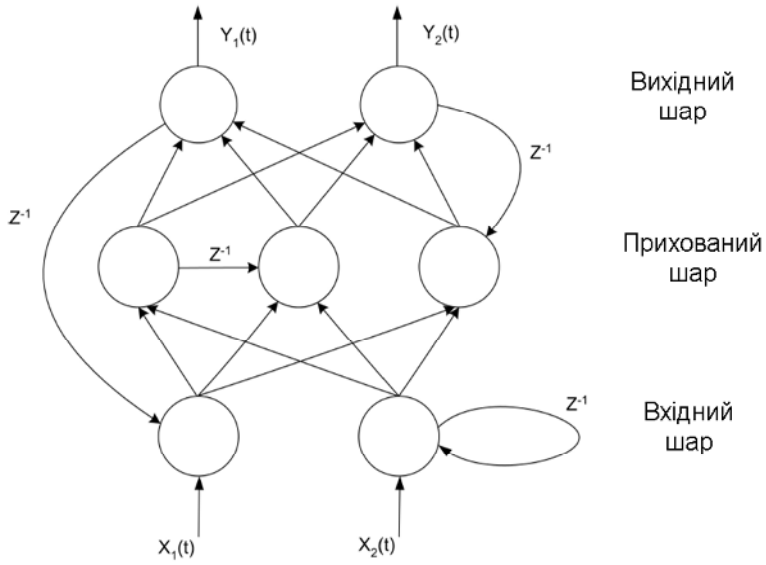

Рис. 3.4. Схема рекурентної нейронної мережі

Однак після численних досліджень було запропоновано кілька модифікацій алгоритму ВР, таких як рекурентний ВР, ВР для послідовностей, рекурентне навчання в реальному часі, часозалежний рекурентний ВР алгоритм і, найбільш популярний серед них, часовий ВР алгоритм. Впровадження цих модифікацій при проведенні навчання БП призвело до підвищення ефективності розпізнавання короткочасних акустико-фонетичних одиниць (фонем). Проте, ці покращення лише мінімально покращили розпізнавання тривалих послідовностей акустичних спостережень, необхідних для адекватного представлення лінгвістичних одиниць, таких як слова. Дослідження також виявили ряд значущих недоліків, які заважають робити RNN основною структурою для СРМ. Ці недоліки включають:

– Відсутність механізмів, що адекватно враховують тимчасову варіативність та послідовну природу мовного сигналу.

- Відсутність теоретичних основ для обчислення чи вибору параметрів, що визначають динаміку та топологію ШНМ (ці параметри часто визначаються на розсуд розробника).
- Незважаючи на наявність різноманітних алгоритмів для прискорення процесу навчання, він залишається ресурсномістким процесом.

РОЗДІЛ 4
КОНСТРУКТОРСЬКА ЧАСТИНА

4.1 Огляд гібридної моделі, що об'єднує багатошаровий персептрон та приховану марківську модель (ПММ)

Як було вказано раніше, при використанні прихованих маріквських моделей у формулі (3.11) потрібно провести оцінку ймовірності емісії, де $p(x_n \mid q_l)$, - ймовірності вектора спостереження x_n при теоретично-визначеному стані ПММ q_l. У першій половині 90-х років минулого століття науковець Боурлард та його учні запропонували ідею використання багатошарового персептрона при оцінці ймовірності $p(q_l \mid x_n)$. Зазначена оцінка представляє собою апостеріорну ймовірність стану ПММ при заданому акустичному векторі. Дану ймовірність можна перетворити в емісійну ймовірність згідно з правилом Байєса.

Нехай ϑ_k при *k=1,…,* де *K* – вихідні дані *k-го* нейрона вихідного шару персептрона, тоді ϑ_k може встановити зв’язок з дискретним ПММ станом S_k, об'єднавши множину параметрів БП Θ_{MLP} з множиною параметрів Θ_{HMM}. При використанні для навчання послідовності акустичних векторів $X = \{x_1, x_2, ..., x_N\}$, вхідним вектором для МП буде акустичний вектор x_N з міткою $q_n = S_k$. Таким чином, можна довести, що в БП є достатня кількість прихованих нейронів для наближення функції відображення один на одного вхідного та вихідного векторів, БП не перенавчається та не потрапляє в локальний мінімум після завершення процедури навчання.

Задовільний результат в даному контексті визначається як результат апроксимації або прогнозування багатошарового персептрона розподілу ймовірностей для різних дискретних станів при заданому вхідному векторі.

$$\vartheta_k(x_n, \Theta_{MLP}^{opt}) = p(S_k \mid x_n, \Theta_{HMM}, \tag{4.1}$$

В якому Θ_{MLP}^{opt} - безліч параметрів, отриманих в процесі навчанні БП. Окрім того, в [16] розглядалося простий варіант розширення наведеної моделі з метою використання контекстуальної інформації, де в якості вхідного сигналу для персептрона використовується послідовність з *2c + 1* акустичних векторів $X_{n-c}^{n+c} + \{x_{n-c},...,x_c,...,x_{n+c}\}$ Таким чином, формула (4.1) може бути переписана, розглядаючи це розширення, наступним чином:

$$\vartheta_k(x_n\Theta_{MLP}^{opt}) = p(q_n = S_k \mid X_{n-c}^{n+c}, \Theta_{HMM} \qquad \forall k = 1,..., K\,. \tag{4.2}$$

Це вдосконалення враховує взаємозв’язок між акустичними векторами, дозволяючи подолати обмеження, що пов'язані зі незалежністю векторів спостережень.

$$\vartheta_k(x_n, \Theta_{MLP}^{opt}) = p(q_k^n \mid q_k^{n-1}, \Theta_{HMM} \quad \forall k = 1,..., K\,. \tag{4.3}$$

Вихідний вектор БП є апроксимацією апостеріорної ймовірності $\vartheta_k(x_n, \Theta_{MLP}^{opt})$ та є оцінкою

$$p(q_k \mid x_n) = \frac{p(x_n \mid q_k)p(q_k)}{p(x_n)}\,, \tag{4.4}$$

що приховано поєднує в собі ймовірність емісії $p(x_n \mid q_k)$ і апріорну ймовірність стану ПММ $p(q_k)$. Оскільки ймовірність у (4.6) входить як множник, це дає можливість регулювати апріорну ймовірність стану під час класифікації, не вдаючись до перенавчання персептрона, а також нормувати вихідні ймовірності персептрона відповідно до використаного навчального мовного корпусу даних. Отже, для використання ймовірності $p(q_k)$ в якості ймовірності емісії для ПММ, потрібно розділити вихід персептрона $q_k(x_n)$ на відносну частоту зустрічі стану S_k в навчальній вибірці. Це дає можливість отримати оцінку виразу:

$$\frac{p(x_n \mid q_k)}{p(x_n)} \tag{4.5}$$

При здійсненні процесу розпізнавання масштабуючий член $p(x_n)$ не змінюється для всіх станів і не має впливу на класифікацію.

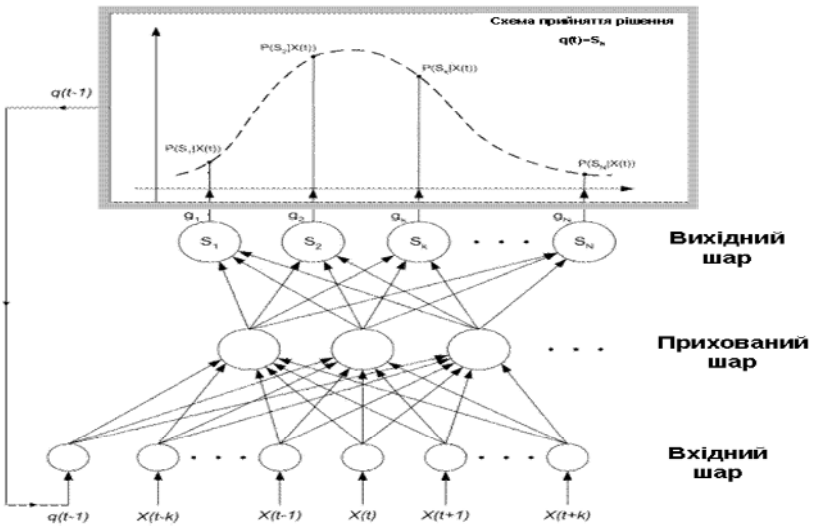

Рис. 4.1 Оцінка ймовірності за допомогою мережі TDNN

4.2 Огляд гібридної моделі, що об'єднує в собі рекурентну мережу та приховану марківську модель

Науковець Робінсон та ін. в праці [18] запропонував модель використання рекурентної мережі замість нейронної мережі із затримкою часу для оцінки ймовірностей емісії ПММ. В основних рівняннях матричні оператори було замінено нелінійною мережею зворотних зв'язків, що призвело до створення нової обчислювальної структури. Поточний акустичний вектор та поточний вектор стану u_n подаються на вхід мережі разом. Ці вектори проходять через типову мережу без зворотних зв'язків з метою отримання вихідного вектора ϑ_n та наступного вектора стану u_n+1. У цьому процесі використовуються матриці ваг зв'язків мережі, позначені як *W* для акустичної мережі та *V* для стану мережі, вхідний вектор z_n, а

$$z_n = \begin{bmatrix} 1 \\ x_n \\ u_n \end{bmatrix}, \tag{4.6}$$

$$\vartheta_n^k = \frac{\exp(W_k z_n)}{\sum_j \exp W_j z_n}, \tag{4.7}$$

$$u_{n+1}^k = \frac{1}{1+\exp - V_k z_n}. \tag{4.8}$$

Введення (4.6) дозволяє здійснити зсув для надання не лінійності стану системі. Аналогічно наведеній моделі Боурларда для використання мережі із затримкою часу, вихід рекурентної мережі представляє собою апостеріорну оцінку ймовірності ПММ стану q_k^n у момент часу n:

$$q_k^n = P(q_k^n \mid X_1^n, u_0) \tag{4.9}$$

Основи такого тлумачення представлені в праці [19]. Використання рекурентних мереж для оцінки ймовірності емісії в гібридних моделях дозволяє досягти значного акустичного контексту, використовуючи вектор внутрішнього стану u_n.

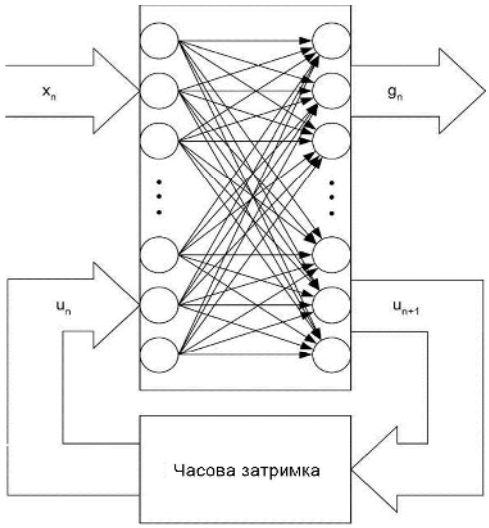

Рис. 4.2 Схема рекурентної нейронної мережі

Отже, подвійний підхід, що об'єднує методи ПММ і ШНМ, взаємодоповнює їхні переваги та компенсує недоліки. Гібридна модель успішно поєднує можливості моделювання довготривалих взаємозалежностей від ПММ та універсальну непараметричну апроксимацію, оцінку ймовірності та дискримінантні алгоритми навчання від ШНМ. Це призводить до зниження кількості параметрів при оцінці та приводить до суттєвого підвищення точності розпізнавання у порівнянні зі традиційними методами.

РОЗДІЛ 5
СПЕЦІАЛЬНА ЧАСТИНА

5.1 Процес створення системи розпізнавання мовних образів

Оптимальним методом створення систем розпізнавання голосових команд є застосування ієрархічного принципу побудови. Кожен модуль такої системи спроектований для розпізнавання образів мови, використовуючи фонемні системи на відповідних рівнів. Іншими словами, на виході кожного модуля отримується набір кандидатів на розпізнавання з найвищою ймовірністю. Логічно припускати, що при переході від модуля більш високого рівня до модуля нижчого рівня граматична складність системи повинна зменшуватися (на початку багато кандидатів на розпізнавання, на кінці - лише одна розпізнана фраза). У загальному вигляді модульна система розпізнавання команд голосу включає такі основні компоненти (рис. 5.1):

- система акустичного введення інформації;
- блок виділення корисного сигналу від шумів навколишнього середовища;
- перед процесор (попередньої обробки);
- блоки розпізнавання - пошуку, що формують ієрархічну структуру;
- блок прийняття та оцінки рішення про розпізнавання.

Кількість необхідних рівнів у ієрархії системи розпізнавання може відрізнятися і встановлюється для кожного конкретного випадку окремо, враховуючи технічні умови експлуатації, наявні фонетичні бази та граматичну складність граматик, а також доступні фінансові ресурси.

Методика проектування та оптимізації СРМ може бути розкрита через послідовне вирішення таких завдань:

1. Проведення аналізу структури системи, що включає вибір акустичної системи введення і формалізацію опису дискурсії системи розпізнавання.

2. Визначення оптимальної конфігурації ієрархічних розпізнавальних модулів на основі застосування теорії побудови оптимальної стратегії розпізнавання. Для

кожного модуля визначають класи звукотипів, які підлягають розпізнаванню, вибирають математичну модель обробки сигналу мови та тип автоматичного класифікатора. На цьому етапі також встановлюються зв'язки різних модулів розпізнавання та їх взаємодія з блоком прийняття і оцінки рішення.

3. Створення математичної моделі процесу розпізнавання у формі направленого графа. Графи можна конструювати як для окремих модулів, так і для об'єднання кількох модулів розпізнавання на одному графі, описуючи їх на різних фонетичних рівнях.

4. Визначення параметрів, що сприяють підвищенню ефективності системи розпізнавання мови: встановлення порогових значень параметрів якості розпізнавання для модулів, де використовуються ПММ, а також налаштування параметрів оцінки достовірності прийнятого рішення [18].

5. Розрахунок аналітичної форми функції динамічного скорочення варіантів перебору при здійсненні процесу розпізнавання [5].

Рис. 5.1. Функціональна схема системи розпізнавання мови

Лексична модель - модель відтворення слова.

Акустична модель - встановлює, які звуки відповідають заданій послідовності слів.

Мовна модель - забезпечує формування послідовності слів.

Важливим етапом у створенні СРМ є вибір методів для виділення корисного сигналу. Проблема відокремлення мови від оточуючого шуму є вкрай складною, за винятком випадків дуже високого співвідношенням сигнал/шум, наприклад, при використанні високоякісних записів у звукоізольованих камерах чи звуконепроникних приміщеннях. У таких умовах енергія навіть слабких звуків мови, таких як фрикативні приголосні, перевищує енергію шуму, і тому достатньо просто виміряти енергію сигналу. Але у більшості реальних ситуацій подібні умови запису зустрічаються рідко.

Розв'язання завдання виділення мови з аудіосигналу є першим кроком у впровадженні фонетичного розпізнавання мови. Цей процес дозволяє вилучити зайвий сигнал як носія інформації. Для вирішення цього завдання пропонується застосовувати такі ознаки сигналу, як його енергія *(E)* і кількість нульових перетинів *(ZC)* для кожного інтервалу фонового шуму.

Для реалізації цього алгоритму потрібно мати статистику шумів, яку можна отримати шляхом вирізання інтервалів шуму з різних аудіосигналів.

5.2 Процес навчання нейронних мереж з вчителем і без вчителя

Характерною особливістю штучних нейронних мереж є їх здатність до процесу навчання, яке включає в себе формування відповідного реагування на різноманітні вхідні сигнали. Навчання нейронних мереж може відбуватися за допомогою різних методів, наприклад:

– зміни конфігурації мережі за допомогою додавання нових зв'язків або вилучення існуючих зв’язків між нейронами;

– модифікації між нейронами одиниць матриці зв'язку (ваг);

– зміни ознак нейронів, таких як форма та параметри активаційної функції, і так далі.

Загальна структура навчання нейронної мережі ілюструється на рис. 5.2.

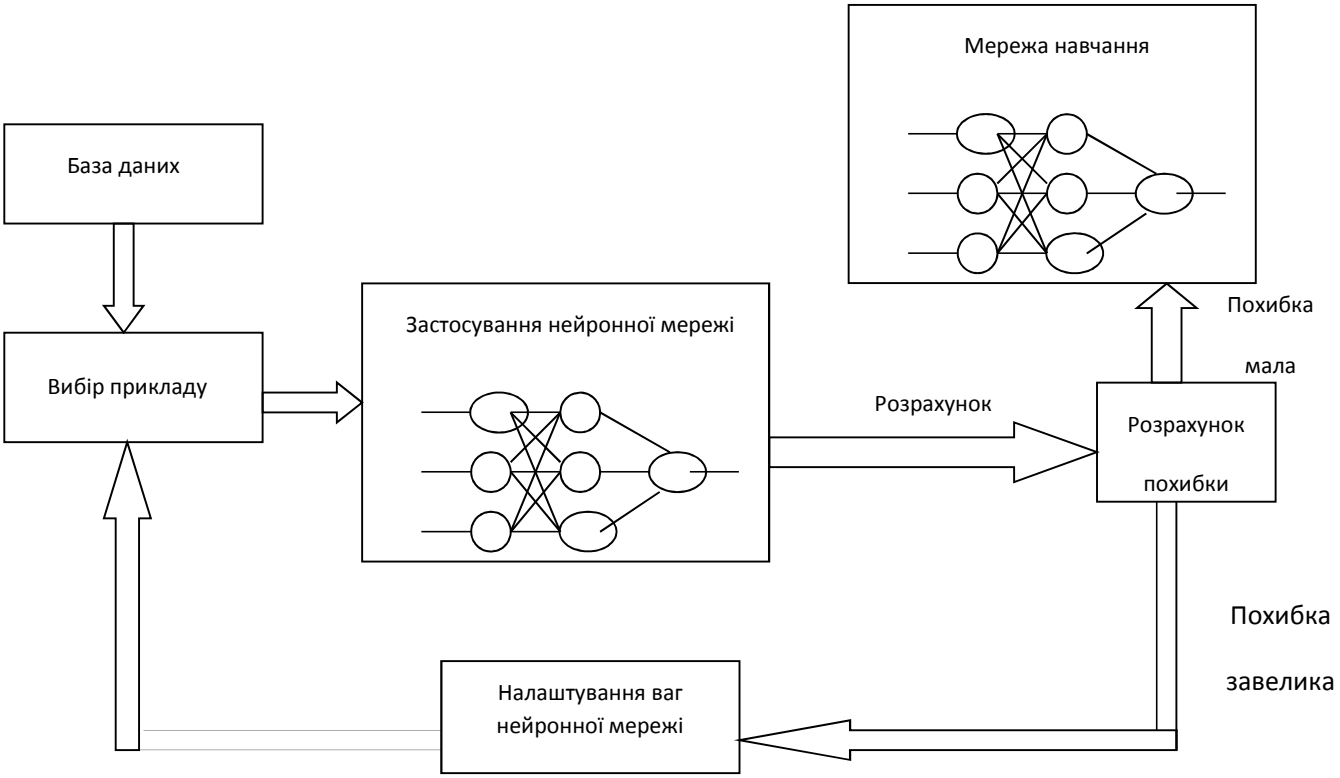

Рис. 5.2 Структура процесу навчання нейронної мережі

Алгоритми навчання можна поділити на дві категорії: процес навчання з вчителем і без. У випадку навчання з вчителем присутній наставник, який подає вхідні дані мережі, порівнює отримані виходи з очікуваними та здійснює налаштування ваг мережі для зменшення розбіжностей. Цей метод, хоч і ефективний для вирішення практичних задач, визнається деякими дослідниками неприйнятним для моделювання біологічних систем, оскільки не відображає механізми, притаманні людському мозку.

У випадку без вчителя мережа самоорганізовується, налаштовуючи свої ваги за певним алгоритмом без наявності конкретних вказівок щодо очікуваного виходу. Внаслідок цього виходи для конкретних нейронів стають непередбачуваними. Тим не менше, мережа організується так, що відтворює важливі характеристики навчального набору. У зв'язку з вибором завдання класифікації образів в даній роботі, був вибраний метод з вчителем, оскільки він добре справляється з задачами класифікації. В табл. 5.1 представлені різні методи алгоритмів навчання з вчителем.

Таблиця 5.1 Алгортм процесу навчання зі вчителем

Навчаюче правило	Архітектура мережі	Алгоритм навчання	Задачі
Корекція похибки	Одношаровий та багатошаровий персепрон	Алгоритм навчання персепрона, зворотнє поширення похибки	Класифікація образів, апроксимація функцій, передбачення керування
Больцман	Рекурентна	Алгоритм навчання Больцмана	Класифікація образів
Хебб	Багатошарова прямого поширення	Лінійний дискримінантний аналіз	Аналіз даних, класифікатор образів
Змагання	Змагання	Векторне квантування	Категоризація всередині класу, стиснення даних
Корекція похибки	Із затримкою часу TDNN	Зворотнє поширення похибки	Класифікація образів

Метод навчання Хеба. Метод навчання Хеба описаний у роботі [1], визначив основні засади більшості алгоритмів навчання нейронних мереж, розроблених після його публікації. Ранні дослідження вказували на те, що процес навчання в галузі біологічних систем зумовлене фізичними змінами в нейронах, але не надавали конкретної вказівки про те, яким чином це відбувається. Враховуючи фізіологічні та психологічні вивчення, Хеб запропонував гіпотезу про те яким чином може відбуватися навчання біологічних нейронів. Його теорія передбачала лише точковий взаємозв’язок між нейронами без участі учителя, тобто навчання відбувається без конкретного керівництва. Навіть без математичного аналізу ідеї Хеба стали загальноприйнятими. Робота науковця стала класичною та широко вивчається дослідниками досі.

Алгоритм процесу навчання Хеба. Хеб суттєво передбачив те, що міць синаптичного з'єднання двох нейронів підсилюється, коли обидва нейрони перебувають у стані збудження. Це свідчить про підсилення синапсу відповідно до кореляції рівнів збудження об'єднаних нейронів, які пов'язані з цим синапсом. Таким чином, алгоритм навчання Хеба отримав назву "кореляційний алгоритм".

Його суть виражена у наступому рівнянні:

$$wij(t+1) = wij(t) + NETi\ NETj \qquad (5.1)$$

де *wij(t)* визначає силу синапсу від нейрона *i* до нейрона *j* у конкретний момент часу *t*; *NETi* представляє рівень збудження предсинаптичного нейрона; *NETi* відображає рівень збудження постсинаптичного нейрона.

Концепція Хеба вирішує складне питання можливості навчання без вчителя. У методі Хеба навчання є виключно локальним, обмеженим двома нейронами та синапсом, який їх з'єднує. Для розвитку нейронна мережа не потребує глобальної системи зворотного зв'язку.

Хоча використання методу Хеба для навчання нейронних мереж призвело до значних досягнень, цей метод також показав свою обмеженість - деякі типи даних просто не можуть бути ефективно використані для навчання за допомогою цього методу. У зв’язку з цим було розроблено чимало розширень і новаторських покращень, більшість з яких в суттєвій мірі ґрунтується на підходах, визначених Хебом у своїй концепції.

Метод процесу сигнального навчання Хеба. Вихід простого штучного нейрону *NET* представляє собою суму його вхідних значень, що може бути виражено наступним чином:

$$NET_j = \sum_i OUT_i W_{ij}, \tag{5.2}$$

де *NETj* визначає вихідну величину *NET* для нейрона *j*, *OUTi* визначає вихід нейрона *i*, а *wij* представляє собою вагу зв'язку між нейроном *j* та нейроном *i*.

Виявлено, що в даному випадку багатошарова лінійна мережа не перевищує одношарову мережу за потужнішою. Здатність розглянутої мережі може покращитися тільки при введенні нелінійності у передатну функцію нейрона. Мережа, яка використовує метод навчання Хеба та сигмоїдальну функцію активації, називають мережею, що навчається за сигнальним методом Хеба. У цьому випадку рівняння Хеба модифікується таким чином:

$$OUT_i = \frac{1}{1+\exp(-NET_i)} = F(NET_i) \tag{5.3}$$

$$wij(t+1) = wij(t) + OUTi\ OUTj \tag{5.4}$$

де *wij(t)* представляє собою силу синапсу від нейрона і до нейрона *j* в конкретний момент часу *t; OUTi* визначає вихідний рівень пресинаптичного нейрона, що дорівнює *F (NETi); OUTj* визначає вихідний рівень постсинаптичного нейрона і дорівнює *F(NET).*

Метод процесу диференційного навчання Хеба відрізняється від методу процесу сигнального навчання тим, що в ньому використовується інша рівність для обчислення зміни ваг. У методі процесу диференційного навчання використовується формула:

$$wij(t+1) = wij(t) + [OUTi(t) - OUTi(t-1)][OUTj(t) - OUTj(t-1)], \quad (5.5)$$

в якому *wij(t)* визначає силу синапсу від нейрона *i* до нейрона *j* у певний момент часу *t*;

OUTi(t) вказує на вихідний рівень пресинаптичного нейрона в певний момент часу *t*;

OUTj(t) визначає рівень постсинаптичного нейрона в момент часу *t*.

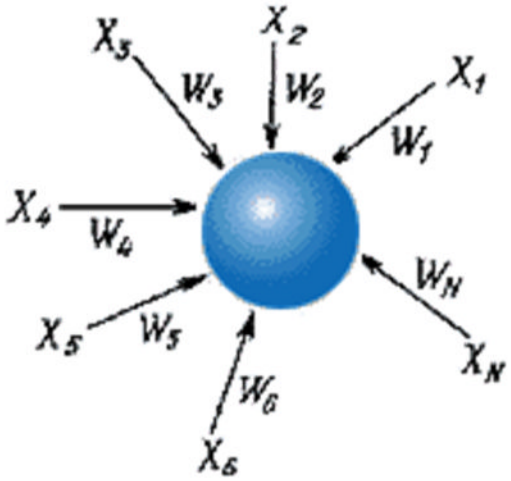

Рис. 5.2. Мережа "Інстар" Гросберга

Вхідні і вихідні зірки. Багато основних концепцій, які використовують в ШНМ, мають своє коріння в роботах Гросберга. Одним із прикладів є конфігурація вхідних і вихідних зірок [21], яка використовується в різних мережевих парадигмах. Вхідна зірка, зображена на рис. 5.2, складається з нейрона, якому подається група входів через синаптичні ваги. Вихідна зірка, яка представлена на рис. 5.3, є нейроном, що контролює групу ваг. Вхідні та вихідні зірки взаємопов'язані в

мережах різного рівня складності. Гросберг розглядає їх як модель певних біологічних функцій. Вигляд зірки визначає її назву.

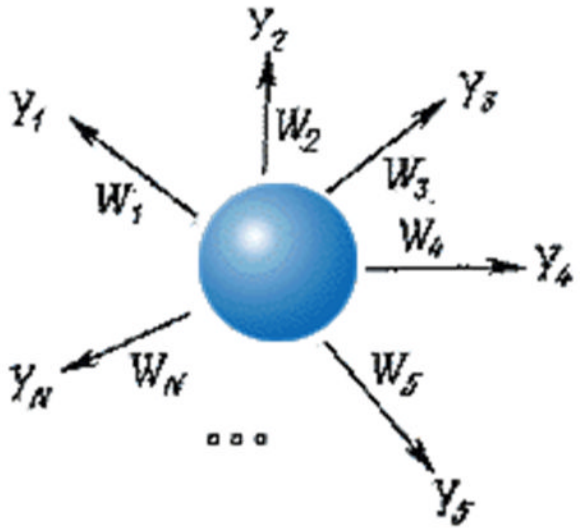

Рис. 5.3. Мережа "Аутстар" Гросберга

Процес навчання вхідної зірки. Вхідна зірка реалізовує процес розпізнавання образів, реагуючи лише на визначений вектор входу *X* і ігноруючи будь-який інший, налаштовуючи ваги, відповідно до *X* -вектора. Вихід визначається як виважена сума входів.

Процес навчання відбувається за формулою

$$wi(t+1) = wi(t) + a\,[xi - wi(t)], \qquad (5.6)$$

де *wi* – вага входу, *xi* - вхід; *a* - нормуючий коефіцієнт навчання з початковим значенням *0,1*, що поступово зменшується.

Після навчання вхідний нейрон реагує на подання вхідного вектора *X*. Навчена зірка може узагальнювати реакції на вектори, адаптуючи ваги під час навчання для усереднення характеристик.

Процес навчання вихідної зірки. У той час як вхідна зірка збуджується кожний раз при появі певного вхідного вектора, вихідна зірка має додаткову функцію; вона виробляє необхідний збуджуючий сигнал для інших нейронів кожний раз, коли збуджується.

Для навчання нейрона вихідної зірки ваги його налаштовуються відповідно до бажаного цільового вектора. Символічний вираз алгоритму навчання можна подати так:

$$wi(t+1) = wi(t) + b\,[yi - wi(t)], \qquad (5.7)$$

де *b* виступає нормуючим коефіцієнтом навчання, який спочатку приблизно дорівнює числу *1* і поступово йде до нуля в процесі здійснення навчання.

Так само, як і у випадку вхідної зірки, ваги вихідної зірки поступово адаптуються під впливом набору векторів так як і звичайні варіації ідеального вектора. У цьому випадку вихідний сигнал нейронів стає статистичною характеристикою навчального набору та, фактично, збігається в процесі навчання до ідеального вектора при пред'явленні лише спотворених версій вектора.

Навчання персептрону. Персептрон представляє собою двошарову, нерекурентну мережу, її структура показана на рисунку 5.4. Вона використовує алгоритм навчання з учителем, що означає, що навчальний набір складається з множини вхідних векторів, при цьому кожен вектор має відповідний вектор-мету. Компоненти вхідного вектора представлені неперервним діапазоном значень, тоді як компоненти вектора-мети є двійковими значеннями *0* або *1*. Після завершення процесу навчання мережа обробляє вхідний набір неперервних даних і формує відповідний вихід у вигляді вектора з двійковими компонентами.

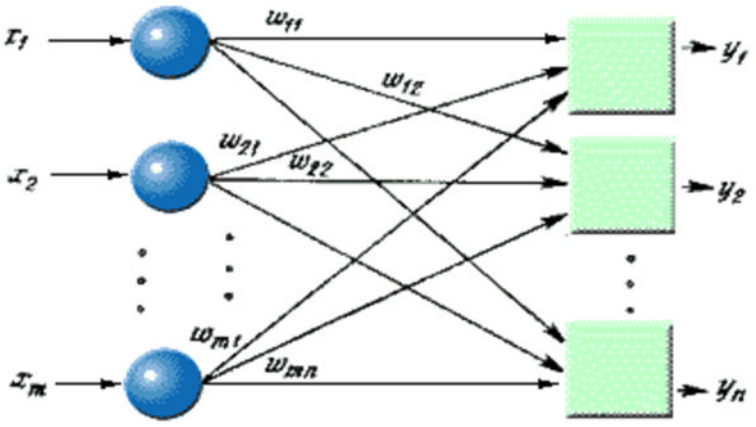

Рис. 5.4. Схема одношарової нейронної мережі

Сам процес навчання здійснюється за наступною процедурою:

Всі ваги мережі початково ініціалізуються невеликими випадковими значеннями.

Вхідний навчальний вектор *X* вводиться у мережу, і сигнал кожного нейрона *NET* обчислюється з використанням стандартного виразу

$$NET_j = \sum_i X_i W_{ij} \quad (5.8)$$

Значення активаційної порогової функції для сигналу NET кожного нейрона розраховується наступним чином:

OUTj = 1, якщо *NETj* більше ніж поріг *θj*,

OUTj = 0 в іншому випадку.

Тут *θ j* - поріг, що відповідає *j* нейрону (в найпростіших випадках, всі нейрони мають один і той самий поріг).

Похибка для кожного нейрона обчислюється шляхом віднімання отриманого виходу від очікуваного виходу:

$$errorj = targetj - OUTj \quad (5.9)$$

Кожна вага підлягає модифікації таким чином:

$$Wij(t+1) = wij(t) + a\ xierrorj \quad (5.10)$$

Кроки від другого до п'ятого повторюються до тих пір, поки значення похибки не стане достатньо малим.

Метод навчання Уидроу-Хофа. Персептрон має обмежені бінарні виходи. Уидроу та Хоф здійснили розширення алгоритму навчання перцептрону, дозволивши йому працювати з безперервними виходами, шляхом використання сигмоїдальної функції [5,6]. Крім цього представили математичний доказ що за окремих умов мережа буде збігатися до любої функції, яку вона може представити. Їхня перша модель «Адалін» включила один вихідний нейрон, а більш пізня модель «Мадалін» розширила цю концепцію у випадках з кількома нейронами виходу.

Вирази, що описують процедуру навчання моделі Адаліна, великою мірою аналогічні тим, що використовуються в персептроні. Основна різниця полягає в четвертому кроці, де використовуються неперервні сигнали *NET* замість бінарних сигналів *OUT*. Модифікований 4 крок в цьому випадку виглядає так:

Похибка для кожного нейрона обчислюється шляхом віднімання отриманого виходу від очікуваного виходу:

5.3 Алгоритм пошуку

При розв'язання задачі аналізу послідовності (АРР) застосовуються різні алгоритми для пошуку параметрів ПММ. До цих алгоритмів відносяться:

- Алгоритм Вітербі,
- А-алгоритм,
- Алгоритм швидкого зіставлення,
- Алгоритм двонаправленого пошуку та інші.

Алгоритм Вітербі використовує пряму ймовірність, що позначається як

$$a_t(j) = P(X_1^t, q_t = S_j \mid \lambda), \tag{5.28}$$

яка є спільною ймовірністю того, що до моменту часу t модель λ перебуватиме в стані S_j, і на виході буде сформована послідовність $X_1^t = x_1 x_2 .. x_1$.

Враховуючи властивості ПММ, $a_t(j)$ визначається так:

$$a_t(j) = \sum_{j=1}^{n-1} P(x_t \mid q_t = S_t) P(q_t = S_t \mid q_{t-1} = S_t) P(X_1^{t-1}, q_{t-1} = S_i \mid \lambda). \tag{5.29}$$

Так як $0<j<N$ і $1<t\le T$, то $a_t(j)$ можна уявити у вигляді рекурентного виразу

$$a_t(j) = \left[\sum_{j=1}^{n-1} a_{t-1}(i) a_{ij} \right] \cdot b_j(x_t), \quad 0 < j < N \;\; i \;\; 1 < t \le T \tag{5.30}$$

з навчальними вимогами $a_1(0) = 1$ і $a_1(j) = a_{0j} b_j(x_1)$, $0<j<N$. З урахуванням кінцевого стану S_n ймовірність формування всієї послідовності X моделлю λ визначається із співвідношення

$$P(X \mid \lambda) = \sum_{i=1}^{N-1} a_t(i) a_{iN} \; . \tag{5.31}$$

Вирази (5.30) і (5.31) дозволяють розрахувати повну ймовірність формування послідовності X в моделі λ з урахуванням усіх доступних можливих шляхів (напрямків зміни станів). Обчислення повної ймовірності $P(X \mid \lambda)$ за формулою

(5.31) вимагає аналізу лише N^2T варіантів для ПММ, в яких із кожного стану дозволено N переходів (кожен стан взаємопов'язаний з кожним).

Алгоритм Вітербі спрямований на знаходження оптимального шляху в моделі λ - послідовності станів, яка найвлучніше відповідає заданій залежності X. На кожному кроці алгоритму запам'ятовуються стани, що забезпечують максимальну спільну ймовірність $P(X_1^t, Q_1^t \mid \lambda)$.. Алгоритм Вітербі ефективно розв'язує задачу відновлення послідовності станів Q за спостережуваною послідовністю символів X.

Позначимо

$$\delta_t(j) = \max_{\forall Q_1^t (q_t = S_j)} P(X_1^t, Q_1^t \mid \lambda) \tag{5.32}$$

де $\delta_t(j)$ ймовірність формування послідовності X_1^t при русі вздовж оптимального шляху, що у момент часу t завершується $q_t = S_j$ станом. За аналогією з прямою ймовірністю, яка визначається рекурентним співвідношенням (5.30), перепишемо $\delta_t(j)$ у вигляді [21]:

$$\delta_t(j) = \max_i \left[\delta_{t-1}(i) a_{ij}\right] b_j(t_t). \tag{5.33}$$

Для відновлення оптимального шляху оптимальної послідовності станів на кожному кроці алгоритму Вітербі зберігає зворотні показники для відповідних оптимальних станів S_j. Масив цих показників позначається як $\psi_t(j)$. Опис алгоритму Вітербі можна подати наступним чином [22]:

1) ініціалізація

$$\begin{aligned} &\delta_t(j) = a_{0j} b_j(x_1), \\ &\psi_1(j) = 0 \end{aligned} \qquad \text{де } 0<j<N, \tag{5.34}$$

2) ітерація

$$\delta_t(j) = \max_{0<t<N} \left[\delta_{t-1}(i) a_{ij}\right] b_j(t_t), \tag{5.35}$$

$$\psi_1(j) = \arg\max_{0<t<N} \left[\delta_{t-1}(i) a_{ij}\right], \tag{5.36}$$

$$1<t<T,\ 0<j<N; \tag{5.37}$$

3) визначення кінцевого стану N

$$\delta_t(N) = \max_{0<t<N}\left[\delta_t(i)a_{iN}\right], \tag{5.38}$$

$$\psi_1(N) = \arg\max_{0<t<N}\left[\delta_{t-1}(i)a_{in}\right] \tag{5.39}$$

4) побудова зворотного шляху

$$q_T^* = \psi_T(N), \tag{5.40}$$

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \text{ де } t=T-1, t-2,\ldots,1, \tag{5.41}$$

де q_t^* - стан ПММ, що належить оптимальному шляху. Алгоритм Вітербі являє собою різновид алгоритму динамічного програмування.

5.4 Процес навчання гібридних моделей

Навчання гібридних моделей представляє собою важливий етап в розвитку сучасних методів обробки сигналів та розпізнавання мови. Цей процес включає в себе інтеграцію прихованих марківських моделей та нейронних мереж для ефективного розв'язання завдань з обробки акустичних даних.

Навчання гібридних моделей є перехресним шляхом між традиційними методами ПММ та передовими досягненнями в області нейронних мереж. Це об'єднання дозволяє покращити точність та ефективність систем розпізнавання мови та обробки акустичних сигналів.

Процес навчання можна загально охарактеризувати за допомогою наступних етапів.

Етап 1. Початкове навчання.

1. Підготовка даних: Початковий етап включає збір та підготовку акустичних даних для навчання моделей. Ручна розмітка та створення моделей тривалості дозволяють підготувати навчальний набір.

2. Моделювання ПММ. Визначення початкових параметрів ПММ для призначення початкових міток. Це створює фреймворк для наступного навчання.

3. Навчання нейронної мережі. Використання нейронної мережі для адаптації до акустичних даних на основі ручної розмітки.

Етап 2. Ітеративний процес навчання.

4. Перерозмітка ітеративним методом: Застосування алгоритму Вітербі для розмітки та перерозмітки даних. Ітеративний підхід дозволяє поступово покращувати точність.

5. Повторне навчання. Використання отриманих результатів для повторного навчання нейронної мережі. Цей процес повторюється для збільшення адаптивності моделі.

Етап 3. Максимізація ймовірностей

6. Максимізація емісійних ймовірностей. Використання градієнтного спуску для максимізації ймовірностей емісії в нейронній мережі.

7. Максимізація транзитивних ймовірностей. Переоцінювання транзитивних ймовірностей за допомогою перегляду моделей тривалостей.

При використанні гібридних моделей з рекурентними мережами, навчання гібридної структури проводиться за допомогою алгоритму Вітербі призначеного для оцінки параметрів системи. Описаний алгоритм спрямований на максимізацію логарифма правдоподібності найбільш ймовірної послідовності станів на основі навчальних даних. Перший прохід алгоритму Вітербі використовується для розмітки послідовності векторів параметрів у термінах станів ПММ. Після цього параметри системи адаптуються з метою збільшення правдоподібності цієї послідовності векторів параметрів.

Основні кроки алгоритму Вітербі полягають у розрахунку ймовірностей початкового стану та емісійних ймовірностей для першого символу в послідовності, рекурсивний проході по моделі для кожного символу в послідовності, розрахунку ймовірностей потрапляння в кожен стан в поточний момент часу, обчисленні найбільш ймовірного шляху, використовуючи інформацію з попереднього кроку. На основі даних кроків відбувається визначення найбільш ймовірного кінцевого стану та побудова найбільш ймовірної послідовності станів, використовуючи збережену інформацію про оптимальні шляхи.

РОЗДІЛ 6
БЕЗПЕКА ЖИТТЄДІЯЛБНОСТІ ТА ОСНОВИ ОХОРОНИ ПРАЦІ

6.1 Охорона праці

Професії, пов'язані з необхідністю тривалої роботи на комп'ютері, характеризуються наявністю різноманітних шкідливих та небезпечних факторів, таких як недостатнє чи надмірне освітлення, умови мікроклімату, рівень шуму, випромінювання та фізичне навантаження [21].

У даному розділі роботи розглядаються вимоги до безпеки організації робочого простору, де проводилися дослідження СРМ. Зокрема, розглядається система освітлення, аналізуються шкідливі та небезпечні фактори, визначається мікроклімат робочої зони, розглядається вплив шуму та засоби захисту від нього, визначається небезпека підвищеного рівня напруженості електромагнітного поля.

Робочі місця в приміщенні розташовані на відстані 1 м одне від одного, у два ряди вздовж стін. Площа одного робочого місця становить 6,0 квадратних метрів, а об'єм - 20,0 кубічних метрів. Загальна площа приміщення складає 36 квадратних метрів, а загальний об'єм - 126 кубічних метрів. Висота стелі приблизно три з половиною метри.

Робочі столи виготовлені із ДСП і мають розміри два метри у довжину та один метр у ширину. Сидіння робочих місць відповідають ергономічним вимогам для забезпечення максимального комфорту та оптимального положення тіла при здійсненні роботи з візуально-дисплейними терміналами (ВДТ). Відстань між користувачем та ВДТ складає близько 700 мм, а висота поверхні робочого столу становить 750 мм.

Під час роботи з ПК виявлено ряд шкідливих та небезпечних факторів, які можуть негативно впливати на здоров'я працівників на робочих місцях, а саме:

- Несприятливі мікрокліматичні умови в робочій зоні;
- Зорові навантаження та напруга;
- Інтелектуальні навантаження;

- Монотонність праці;
- Нервово-емоційна напруга;
- Невідповідність ергономічних параметрів робочого місця чинним вимогам;
- Фізична важкість роботи та статичні навантаження на м'язово-скелетний апарат;
- Шум та його вплив;
- Електромагнітне випромінювання;
- Ризики електричного ураження;
- Ризики пожежі;
- Недостатня або занадто висока освітленість робочого місця;
- Ризик виникнення надзвичайних ситуацій природного чи штучного характеру на об'єкті або території.

Електромагнітні поля, що характеризуються напруженням електричних та магнітних компонентів, представляють найбільший ризик для здоров'я людини. Центральним джерелом цих проблем для осіб, які використовують автоматизовані інформаційні системи на базі персональних комп'ютерів, є екрани (монітори).

Портативні комп'ютери випромінюють різноманітні типи випромінювань, включаючи:

- М'яке рентгенівське випромінювання;
- Ультрафіолетове випромінювання (в діапазоні 200-400 нм);
- Видиме світло (в діапазоні 400-700 нм);
- Ближньоінфрачервоне випромінювання (в діапазоні 700-1050 нм);
- Радіочастотне випромінювання (від 3 кГц до 30 МГц);
- Електростатичні поля.

У великих дозах ультрафіолетове випромінювання може викликати дерматит, головний біль та подразнення очей. Інфрачервоне випромінювання може призвести до перегріву тканин, особливо хрусталика ока, і підвищення температури тіла. Напруженість електростатичних полів повинна утримуватися на рівні не більше 20 кВ/м, а поверхневий електростатичний потенціал не повинен перевищувати 500 В. [21].

При підвищеному рівні напруженості полів рекомендується обмежувати період роботи за комп'ютером та робити короткі перерви протягом кожних півтори години, використовувати захисні екрани. Заземлені захисні екрани, зроблені з дрібної сітки або скла, приймають на себе електростатичний заряд, який заземлюється для його зняття.

Для контролю рівня мікроклімату використовуються Санітарні норми мікроклімату виробничих приміщень (ДСН 3.3.6.042-99). Нормування параметрів мікроклімату здійснюється в залежності від сезону та характеру виконуваних робіт. Для стійких робочих місць, зокрема для користувачів ПК, законодавчо встановлені оптимальні параметри мікроклімату. У випадку неможливості їх забезпечення слід використовувати допустимі значення. Операції операторів ПК віднесені до категорій легких робіт Іа, Іб. У таблиці 6.1 представлені оптимальні значення параметрів мікроклімату для приміщень, де виконуються операції операторського характеру [10].

Таблиця 6.1 Параметри мікроклімату для приміщень з ПК

Період року	Параметр мікроклімату	Величина
Холодний	Температура повітря в приміщенні Відносна вологість Швидкість руху повітря	22...24°С 40... 60% до 0,1 м/с
Теплий	Температура повітря в приміщенні Відносна вологість Швидкість руху повітря	23...25 °С 40…60% 0,1...0,2 м/с

Виміряні показники температури та вологості у приміщенні відповідають значенням, вказаним у таблиці для теплого та холодного періоду року. Важливо відзначити, що для нормалізації параметрів мікроклімату слід також використовувати системи кондиціювання повітря або забезпечити подачу свіжого повітря за допомогою вентиляційних систем. Нормативи подачі свіжого повітря наведені у таблиці 6.2. [10].

Таблиця 6.2 Норми подачі свіжого повітря в приміщення з ПК

Характеристика приміщення	Об'ємна витрата свіжого повітря, що подається в приміщення, $м^3$ на одну людину в
Об'єм до 20$м^3$ на людину 20... 40 $м^3$ на людину Більше 40 $м^3$на людину	Не менше 30 Не менше 20 Може біти використана природна вентиляція

Для уникнення перевтомлення необхідно виконувати вправи для очей та дотримуватися оптимального графіку роботи та відпочинку. На робочому місці було впроваджено режим відпочинку, забезпечуючи перерви кожні дві години для виконання фізичних вправ, спрямованих на розслаблення м'язів очей.

Для освітлення приміщення, в якому проводилася дослідження, використовувалися люмінесцентні лампи. У порівнянні з лампами розжарювання вони володіють кількома суттєвими перевагами, такими як спектральний склад, близький до природного світла, підвищена світлова віддача (в 2-5 разів вища, ніж у ламп розжарювання) та триваліший термін служби (до 10 тисяч годин).

У приміщенні використовуються світильники типу ОД, кожен з яких обладнаний двома лампами. Однак, на момент крайньої атестації робочого місця справно працювало лише 12 ламп, замість рекомендованих 20, що призвело до невідповідності рівня штучного освітлення санітарним нормам.

Для покращення робочих умов рекомендується підвищити рівень загальної освітленості приміщення шляхом додаткового встановлення 8 ламп.

В контексті охорони праці, яка стає дедалі більш актуальною, було проведено дослідження небезпечних та шкідливих умов праці на робочому місці. Розроблені рекомендації спрямовані на усунення факторів, що зв’язані зі недостатньою освітленістю, для покращення умов праці та зниження ризику перевтомлення працівників.

6.2 Безпеки в надзвичайних ситуаціях

При використанні комп'ютерів важливо ураховувати різні фактори ризику, включаючи фізичні та психофізіологічні. Забезпечення електробезпеки є ключовим аспектом у цьому контексті. Заходи для уникнення небезпеки від ураження електричним струмом включають правильне розташування обладнання та електричних кабелів. Додаткові заходи для забезпечення електробезпеки включають використання надійних розеток, вогнестійких матеріалів та розрахунок потужності електромережі для витримки підвищеного навантаження.

Для попередження пожежі та поліпшення умов праці рекомендується використовувати надійні розетки, скриту електромережу, встановлення захисних екранів, а також регулярне обслуговування обладнання для зменшення впливу пилу.

В системі освітлення важливо дотримуватися ряду специфікацій, таких як рівень освітленості на робочому місці, рівномірний розподіл яскравості на поверхні монітора, відсутність різких тіней та відблисків, а також стабільність освітленості протягом робочого часу. Рекомендується враховувати оптимальну спрямованість світлового потоку та вибирати необхідний спектр світла.

Основне обладнання користувача комп'ютера на робочому місці включає в себе монітор, клавіатуру та системний блок. Планування розташування робочих місць передбачає відстань не менше 1,5 м. від стін з вікнами, 1 м. від інших стін та не менше 1,5 м. між собою. Оптимально робоче місце розташовувати відносно вікон так, забезпечити спадання природнього світла збоку, переважно зліва, уникаючи прямого попадання світла в очі.

Для забезпечення комфортного освітлення важливо розміщувати джерела світла паралельно напрямку погляду, використовуючи антиполіскові сітки, фільтри для екранів та захисні козирки для уникнення відблисків. Монітор слід розташовувати перпендикулярно до напрямку погляду, забезпечуючи дистанцію від очей, що трохи перевищує звичайну відстань між очима та книгою.

Для запобігання сутулості та забезпечення оптимальних умов перегляду, рекомендується використовувати перед екраном монітора захисний екран,

забезпечуючи також створення неоднорідного поля зору за допомогою картин та плакатів на стінах приміщення.

Під час роботи з текстовою інформацією найбільш фізіологічно-доцільним є використання чорних знаків на світлому фоні, а елементи робочого місця розміщувати так, щоб зберігалася однакова відстань екрана до очей, клавіатури та тексту.

Забезпечення комфортної пози під час роботи з комп'ютером досягається за допомогою регулювання висоти крісла, робочого столу та підставки для ніг. Оптимальна робоча поза враховує горизонтальне розташування ступнів працівника, горизонтальну орієнтацію стегон, вертикальне положення верхніх частин рук. Кут згину ліктевого суглоба варіюється між 70 і 90°, а кут згину зап'ястя не перевищує 20° при нахилі голови від 15 до 20°.

Суттєвим аспектом є спинка крісла, яка повинна адаптуватися до форми спини користувача, уникати тиску на стегно чи куприк. Висоту крісла слід регулювати так, щоб уникнути дискомфортного тиску. Крім того, крісло можна обладнувати бильцями та розташовувати його так, щоб до клавіатури не потрібно було надто довго тягтися.

Для забезпечення здоров'я при тривалій роботі за комп'ютером важливо враховувати регулярні перерви, фізичні вправи та зміну поз. Рекомендована перерва 15 хвилин протягом кожної години та легкі вправи для розслаблення м’язів кілька разів на годину.

Щоб уникнути негативного впливу статичної електрики, рекомендовано підвищувати вологість повітря у приміщенні, де використовуються комп'ютери, через кімнатні зволожувачі. Рекомендується уникати використання синтетичного одягу для запобігання комп'ютерних захворювань, таких як хвороби серцево-судинної, шлунково-кишкової систем, захворювання органів зору.

Перед початком роботи важливо виконати наступні вимоги безпеки:

- Активувати систему кондиціювання у приміщенні.
- Перевірити надійність фіксації обладнання на робочому столі, оптимально налаштувати монітор для забезпечення комфортного кута огляду - під

прямим кутом, трохи зверху вниз і з нахиленим екраном, де нижній край знаходиться ближче до оператора.

- Перевірити стан загальної апаратури: електропроводи, з'єднувальні шнури, штепсельні вилки, розетки, правильність заземлення захисного екрана.
- Адаптувати освітленість робочого місця до комфортного рівня.
- Налаштувати і зафіксувати висоту крісла з зручним нахилом його спинки.
- Підключити необхідну апаратуру до системного блоку, дотримуючись правила вставляти та виймати кабелі при вимкненому комп'ютері.
- Увімкнути апаратуру комп'ютера поетапно: спочатку монітор, потім системний блок, а також принтер, якщо планується його використання.
- Налаштувати яскравість, мінімальний розмір світлої точки, контрастність монітора та фокусування, уникати надто яскравого зображення для збереження зіру та убезпечення очей від втоми.

Вимоги безпеки під час виконання роботи:

- Клавіатуру слід стійко розташовувати на робочому столі, уникати її хитання, сидіти прямо та утримувати розслаблену позу під час роботи на клавіатурі.
- Для пристроїв типу "миша" слід забезпечити велику вільну поверхню столу для їх переміщення та забезпечити зручний упор для ліктьового суглоба, щоб уникнути негативного впливу на користувача.
- Важливо утримувати робоче середовище вільним від посторонніх розмов та подразнюючих шумів.
- Періодично, при вимкненні комп'ютера, слід очищати апаратуру від пилу за допомогою ледь вологої мильної ганчірки. Екрани рекомендується протирати ганчіркою, змоченою у спирті. Заборонено використання рідин та аерозольних засобів-чищення поверхонь комп'ютера.

Заборонено:

- Розміщувати будь-які предмети на апаратурі комп'ютера .

- Закривати вентиляційні отвори апаратури, оскільки це може призвести до перегрівання та виходу з ладу обладнання.

З метою зняття статичної електрики рекомендується періодично торкатися металевих поверхонь.

У цьому розділі розглянуті питання впровадження різновидів організації робочого місця з метою забезпечення безпеки та попередження негативних впливів ПК на здоров'я користувача.

ВИСНОВКИ

У магістерській кваліфікаційній роботі було ретельно розглянуто різноманітні підходи з метою вирішення актуальної проблеми, а саме: розробка системи автоматизованого розпізнавання мови, заснованої на гібридній моделі, що об'єднує приховані марківські моделі та штучні нейронні мережі. Основні результати проведених досліджень можна узагальнити наступним чином:

1. У рамках магістерського проекту було проведено аналіз розробки голосових інтерфейсів, що ґрунтується на механізмі розпізнавання фраз користувача. Цей механізм базується на математичних моделях аналізу та класифікації голосових команд
2. Здійснено ретельний аналіз літературних джерел, що стосуються тематики магістерської роботи.
3. Розглянута гібридна модель комбінованої мови, яка представляє собою синтез прихованих марківських моделей та штучних нейронних мереж, що дозволяє ефективно поєднувати плюси прихованих марківських моделей із можливостями штучних нейронних мереж. Зокрема, ПММ забезпечує моделювання тривалих залежностей, тоді як ШНМ надає непараметричну універсальну апроксимацію, оцінку ймовірностей, алгоритми дискримінантного навчання та зменшення кількості параметрів, необхідних для оцінки, які зазвичай є необхідними для звичайних прихованих марківських моделей.
4. Запропоновано вибір для гібридної моделі нейронних мереж з рекурентною архітектурою та мереж із затримкою часу.
5. Розглянуто алгоритм навчання для запропонованих нейронних мереж, а саме: алгоритм Больцмана для навчання рекурентних мереж та алгоритм зворотного поширення похибки для мереж із затримкою часу.
6. Запропонований алгоритм Вітербі для пошуку параметрів прихованих марківських моделей при вирішенні задачі автоматизованого розпізнавання мови.

7. Розроблена функціональна схема системи автоматизованого розпізнавання мови.

Моїм особистим внеском у кваліфікаційну магістерську роботу на тему "Розробка автоматизованої системи розпізнавання мови з використанням прихованих марківських моделей та нейронних мереж" є систематизація та аналіз передового наукового досвіду в галузі мовного розпізнавання, а також впровадження нових ідей та підходів у даному контексті.

По-перше, я розглядав та оцінював переваги та обмеження прихованих марківських моделей та нейронних мереж у контексті розпізнавання мови. Мій аналіз дозволив виявити, що комбінування обох підходів може суттєво покращити точність та ефективність системи.

Другий важливий аспект мого внеску - це розробка та вдосконалення алгоритмів обробки сигналів для підвищення якості вхідних даних. Це має вирішальне значення для точності розпізнавання мови, особливо в умовах шуму та інших спотворень.

Третій аспект мого внеску полягає в застосуванні глибокого навчання для покращення ефективності нейронних мереж у завданні розпізнавання мови. Використання важливих архітектур, таких як рекурентні та згорткові нейронні мережі, а також використання відомих функцій активації, дозволило досягти значущих результатів у забезпеченні високої точності та швидкості розпізнавання.

Мій особистий внесок полягає також у розробці та оптимізації програмного забезпечення для реалізації запропонованої системи. Відповідно до висновків та рекомендацій, отриманих під час дослідження, програмне забезпечення враховує всі найсучасніші підходи та методи, що гарантує стабільну та ефективну роботу системи.

Усі ці складові мого внеску об'єднуються з метою створення автоматизованої СРМ, яка відповідає найвищим стандартам точності та продуктивності, що сприяє не лише практичному вдосконаленню технічних аспектів системи, але й вносить свій внесок у розвиток області мовного розпізнавання в цілому.

Отримані результати дослідження та розробки автоматизованої системи розпізнавання мови з використанням прихованих марківських моделей та нейронних мереж мають значущі практичні застосування в різних сферах. Нижче наведено деякі з практичних вигод та можливостей, які можуть виникнути з використання отриманих результатів:

1. Підвищення точності розпізнавання мови: Запропонована система, комбінуючи приховані марківські моделі та нейронні мережі, може досягти високої точності розпізнавання мови навіть в умовах шуму чи інших спотворень. Це робить систему ефективною в реальних сценаріях використання, таких як розпізнавання мови в шумному оточенні або в системах голосового керування.

2. Широкі можливості застосування в індустрії та бізнесі: Система може бути успішно використана в індустріальних та бізнесових сценаріях, зокрема для автоматизованого оброблення голосових команд, створення диктантів або автоматизованого аналізу телефонних розмов. Це може покращити ефективність роботи та взаємодії з інформацією.

3. Розвиток систем голосового інтерфейсу: Результати дослідження можуть сприяти подальшому розвитку голосових інтерфейсів в різних пристроях, таких як мобільні телефони, планшети, домашні асистенти та інші. Висока точність розпізнавання забезпечить зручність та ефективність користувачів.

4. Медичні застосування: В сфері медицини система може бути використана для розпізнавання та документування мовленнєвих патологій, а також для розвитку інтерфейсів для людей з обмеженими можливостями.

5. Розширення мовних можливостей інтернет-платформ**:** Застосування системи в інтернет-сервісах, таких як пошукові системи, асистенти та інші, може покращити розпізнавання голосових запитань та команд, забезпечуючи більш швидку та точну взаємодію з користувачами.

Отже, отримані результати мають потенційно значущий вплив на різні аспекти сучасного суспільства та можуть сприяти створенню нових, вдосконалених технологій для розпізнавання мови в різних галузях життя.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Грищук Т.В., Биков М.М., Підвищення швидкодії розпізнавання мови прихованими марківськими моделями//Комп'ютерні технології друкарства. – Львів. Українська академія друкарства, 2005. –№13.–С.99-107.

2. Грищук Т.В., Биков М.М., Ієрархічна стратегія розпізнавання мови// Вісник Технологічного університету Поділля. - 2004. - №2. – Т. 2. – С.58-61.

3. Грищук Т.В., Раїмі А.А., Биков М.М., Використання нейронних мереж для розпізнавання звуків мови//Оптикоелектронні інформаційно-енергетичні технології. – 2001. - №2. – С.92-97.

4. Грищук Т.В., Биков М.М., Розробка методів оцінки ефективності автоматизованих систем розпізнавання мови//Вісник Технологічного університету Поділля. – 2003. – №3. Т. 1 - С.122-125.

5. Грищук Т.В. Розпізнавання природної мови на граматичних марківських мережах//Наукові праці Донецького національного технічного університету. Серія: "Обчислювальна техніка та автоматика". – Донецьк: Дон. НТУ, 2005.–С. 181-187.

6. Грищук Т.В., Биков М.М., Оцінка впливу рівня шумів на ефективність систем розпізнавання слів української мови//Матеріали сьомої міжнародної науково-технічної конференції "Контроль і управління в складних системах" (КУСС – 2003). – Вінниця: УНІВЕРСУМ - Вінниця, 2003. – С. 65-70.

7. Грищук Т.В., Биков М.М., Методи підвищення дикторонезалежності опису і розпізнавання мовної інформації в мережі INTERNET//Третя міжнародна конференція "Інтернет-Освіта-Наука-2002" (ІОН – 2002).– Вінниця: УНІВЕРСУМ-Вінниця, 2002.–Т 2–С.329-332.

8. Грищук Т.В., Биков М.М., Розпізнавання мовних образів з використанням нейромережевого підходу//Праці міжнародної конференції з індуктивного моделювання (МКІМ-2002.).–Львів: Державний НДІ інформаційної інфраструктури, 2002.–Т.1., Ч.2.–С.203-207.

9. Миколюк І.О. Аналіз методів розпізнавання мовлення//Вісник Вінницького національного технічного університету – Вінниця, 2020 – С. 201-204

10. Санітарні норми мікроклімату виробничих приміщень ДСН 3.3. 6.042 99 – [Електронний ресурс] – Ресурс доступу: https://zakon.rada.gov.ua/rada/show/va042282-99#Text

11. Zavaliagkos G., Austin S., Schwartz R., Makhoul J. Speech recognition using segmental neural nets//IEEE ICASSP, San Francisco, March - 1992, - pp.I-625-628.

12. Zavaliagkos G., Austin S., Schwartz R., Makhoul J. Improving state-of the-art continuous speech recognition system using the N-best paradigm with neural networks//Proceedings

13. DARPA Speech and Natural Language Workshop, Harriman, NY (Morgan Kaufmann, Los Altos, CA).-1992.-pp.180-184.

14. Wellekens C. Links Between Markov Models and Multilayer Perceptrons//IEEE Transactions on Pattern Analysis and Machine Intelligence. -1990.- Vol. 12. – pp.1167-1178.

15. Murveit H., Weintraub M., Cohen M., Bernstein H., Price P. The decipher speech recognition system//IEEE ICASSP, Albuquerque, - 1990. - pp.77-80.

16. Laird N., Dempster A., Rubin D. Maximum likelihood from incomplete data via the EM algorithm//J. Roy. Stat. Soc. - 1977. - Vol. 39. - pp. 1-38.

17. Cohen M., Franco H., Morgan N., Abrash V., Rumelhart D. Context-dependent connectionist probability estimation in a hybrid hidden Markov model-neural net speech recognition system//Computer Speech and Language. - 1994. - pp. 211- 222.

18. Ris C., Henneberg J., Bourlard H., Morgan N, Renals S. Estimation of global posteriors and forward-backward training of hybrid HMM/ANN systems//Proceedings of EUROSPEECH, - 1997. - pp. 1951-1954.

19. Renals S. J., Hochberg M, Kershaw D., Robinson A. Large vocabulary continuous speech recognition using a hybrid connectionist-HMM system//Proceedings of CSLP, Yokohama. - 1994. - pp. 1499-1502.

20. Таран С.А. Головні проблеми розробки нових систем розпізнавання мови і шляхи їх вирішення роботи // Матеріали XI науково-технічної конференції

«Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя,. – Тернопіль, ТНТУ, 2023. – С.180-181

21. Жидецький В.Ц. Охорона праці користувачів комп’ютерів. Навчальний посібник. – Вид. 2-ге., доп. – Львів.: Афіша, 2000. – 176с.

22. Васильєва Н.Б., Федорин Д.Я. Проблеми створення систем розпізнавання мовлення для різних комп’ютерних платформ // ISSN 1561-5359 «Штучний інтелект». - №4. – Київ, 2013. – С. 158-167

23. Проектування мікропроцесорних систем керування: навчальний посібник/ І.Р. Козбур, П.О. Марущак, В.Р. Медвідь, В.Б. Савків, В.П. Пісьціо.– Тернопіль: Вид-во ТНТУ імені Івана Пулюя, 2022.–324с.

24. Я.І. Проць, В.Б. Савків,О.К. Шкодзінський, О.Л. Ляшук. Автоматизація виробничих процесів. Навчальний посібник для технічних спеціальностей вищих навчальних закладів. – Тернопіль: ТНТУ ім. І.Пулюя, 2011. – 344с.

25. Основи наукових досліджень і теорія експерименту : Навчальний посібник / укл. Ю. Б. Капаціла, П. О. Марущак, В. Б. Савків, О. П. Шовкун. Тернопіль : ФОП Паляниця В.А., 2023. 186 с.». http://elartu.tntu.edu.ua/handle/lib/40843.

26. Пилипець М. І. Правила заповнення основних форм технологічних документів : навч.-метод. посіб. / Уклад. Пилипець М. І., Ткаченко І. Г., Левкович М. Г., Васильків В. В., Радик Д. Л. Тернопіль : ТДТУ, 2009. 108 с. https://elartu.tntu.edu.ua/handle/lib/42995.

27. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної та заочної (дистанційної) форм навчання «Безпека в надзвичайних ситуаціях» / В.С. Стручок –Тернопіль: ФОП Паляниця В. А., –156 с. https://elartu.tntu.edu.ua/handle/lib/39196.

28. Навчальний посібник «Техноекологія та цивільна безпека. Частина «Цивільна безпека»» / автор-укладач В.С. Стручок – Тернопіль: ФОП Паляниця В. А., – 156 с. http://elartu.tntu.edu.ua/handle/lib/39424/

29. Платформа .NET та мова програмування C# 8.0: навчальний посібник / Коноваленко І.В., Марущак П.О. – Тернопіль: ФОП Паляниця В. А., 2020 – 320 с.

/Рекомендовано до друку Вченою радою Тернопільського національного технічного університету імені Івана Пулюя. Протокол № 10 від 20 жовтня 2020 року

30. Савків В.Б., Капаціла Ю.Б., Михайлишин Р.І. Методичні вказівки до виконання кваліфікаційної роботи бакалавра спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології». Тернопіль.: Видавництво ТНТУ. 2021. 50 с. https://elartu.tntu.edu.ua/handle/lib/35172

31. А.Г. Микитишин, М.М. Митник, П.Д. Стухляк, В.В. Пасічник Комп'ютерні мережі. Книга 1. [навчальний посібник] (Лист МОНУ №1/11-8052 від 28.05.12р.) - Львів, "Магнолія 2006", 2013. – 256 с.

32. А.Г. Микитишин, М.М. Митник, П.Д. Стухляк, В.В. Пасічник Комп'ютерні мережі. Книга 2. [навчальний посібник] (Лист МОНУ №1/11-11650 від 16.07.12р.) - Львів, "Магнолія 2006", 2014. – 312 с.

33. Микитишин А.Г., Митник, П.Д. Стухляк. Комплексна безпека інформаційних мережевих систем: навчальний посібник – Тернопіль: Вид-во ТНТУ імені Івана Пулюя, 2016. – 256 с.

34. Микитишин А.Г., Митник М.М., Стухляк П.Д. Телекомунікаційні системи та мережі : навчальний посібник для студентів спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології» – Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя, 2017 – 384 с.

35. Введення в компютерну графіку та дизайн: Навчальний посібник для студентів спеціальності 174 "Автоматизація, компютерно-інтегровані технології та робототехніка"/Укладачі: О.В. Тотосько, П.Д. Стухляк, А.Г. Микитишин, В.В. Левицький, Р.З. Золотий - Тернопіль: ФОП Паляниця В.А., 2023 - 304с. http://elartu.tntu.edu.ua/handle/lib/41166.